

TOWARDS UNDERSTANDING MODE-OF-ACTION OF TRADITIONAL MEDICINES BY USING *IN SILICO* TARGET PREDICTION

Siti Zuraidah Binti Mohamad Zobir
Hughes Hall

April 2017

This dissertation is submitted for the degree of Doctor of Philosophy of the
University of Cambridge



Supervisor:

Dr Andreas Bender

Name: SITI ZURAIDAH BINTI MOHAMAD ZOBIR

Title: *IN SILICO* TARGET PREDICTION: TOWARDS UNDERSTANDING MODE-OF-ACTION OF TRADITIONAL MEDICINES

Abstract

Traditional medicines (TM) have been used for centuries to treat illnesses, but in many cases their modes-of-action (MOAs) remain unclear. Given the increasing data of chemical ingredients of traditional medicines and the availability of large-scale bioactivity data linking chemical structures to activities against protein targets, we are now in a position to propose computational hypotheses for the MOAs using *in silico* target prediction. The MOAs were established from supporting literature. The *in silico* target prediction, which is based on the “Molecular Similarity Principle”, was modelled *via* two models: a Naïve Bayes Classifier and a Random Forest Classifier. Chapter 2 discovered the relationship of 46 traditional Chinese medicine (TCM) therapeutic action subclasses by mapping them into a dendrogram using the predicted targets. Overall, the most frequent top three enriched targets/pathways were immune-related targets such as tyrosine-protein phosphatase non-receptor type 2 (PTPN2) and digestive system such as mineral absorption. Two major protein families, G-protein coupled receptor (GPCR), and protein kinase family contributed to the diversity of the bioactivity space, while digestive system was consistently annotated pathway motif. Chapter 3 compared the chemical and bioactivity space of 97 anti-cancer plants’ compounds of TCM, Ayurveda and Malay traditional medicine. The comparison of the chemical space revealed that benzene, anthraquinone, flavone, sterol, pentacyclic triterpene and cyclohexene were the most frequent scaffolds in those TM. The annotation of the bioactivity space with target classes showed that kinase class was the most significant target class for all groups. From a phylogenetic tree of the anti-cancer plants, only eight pairs of plants were phylogenetically related at either genus, family or order level. Chapter 4 evaluated synergy score of pairwise compound combination of Shexiang Baixin Pill (SBP), a TCM formulation for myocardial infarction. The score was measured from the topological properties, pathway dissimilarity and mean distance of all the predicted targets of a combination on a representative network of the disease. The method found four synergistic combinations, ginsenoside Rb3 and cholic acid, ginsenoside Rb2 and ginsenoside Rb3, ginsenoside Rb3 and 11-hydroxyprogesterone and ginsenoside Rb2 and ginsenoside Rd agreed with the experimental results. The modulation of androgen receptor, epidermal growth factor and caspases were proposed for the synergistic actions. Altogether, *in silico* target prediction was able to discover the bioactivity space of different TMs and elucidate the MOA of multiple formulations and two major health concerns: cancer and myocardial infarction. Hence, understanding the MOA of the traditional medicine could be beneficial in providing testable hypotheses to guide towards finding new molecular entities.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Acknowledgements

I would like to gratefully acknowledge various people who have been helping me throughout my journey of producing this thesis. Firstly, I would like to express my sincerest appreciation to my supervisor, Dr Andreas Bender for the opportunity to be part of his group and all his guidance and deep insights at each step of my study. I've learned extensively from him including how to critically approach a research problem, make a rational decision and present a good research output.

Special thanks should be given to Dr. Tai-Ping Fan for the tremendous help in explaining the concept of traditional Chinese medicine (TCM) and pharmacology. Also, to Dr Xian Jun Fu, a visiting researcher from China, for his input on TCM and Ranjoo Choi for her patience and dedication to run the experiments in the lab to validate TCM compound combinations.

I want to thank past members of the group, Fazlin Mohd Fauzi, Gergios Drakakis and Sonia Linggi for their helps during the challenging first year to adapt to this field of cheminformatics. Many thanks go out to Avid Afzal, Azedine Zoufir and Lewis Mervin for answering all my statistics-related questions. I would also like to thank Deszo Modos for the biological network tutorial sessions and Krishna Bulusu for the thoughtful comments on my research projects. A thanks also goes to Fatima Baldo for proofreading the thesis. Not to forget, to all members in the group, thank you for all the helps I have received. The group has been a source of good advice, diverse leaning experience, and friendships.

I would like to express my appreciation to Susan Begg and Chloe Baker for being so helpful and providing their administrative assistance. I also appreciate the financial support of Majlis Amanah Rakyat (Malaysia) and Malaysian Institute of Pharmaceuticals (IPharm).

Finally, I would like to thank my family for all their love and encouragement. To both my parents, thank you for the annual visit to Cambridge and to all my siblings, thank you for all practical support in all those things beyond doing a PhD. I'm also grateful for many new friends I have made during these four years. Thank you for the fun we had in Cambridge or during our travels to several European countries.

Correspondence to previous publications

One part of this thesis has been published previously in refereed journals. Please find the article corresponding to the chapter of this thesis below.

Chapter 2:

Zobir, S. Z. M.; Fauzi, F. M.; Liggi, S.; Drakakis, G.; Fu, X.; Fan, T.-P.; Bender, A., Global Mapping of Traditional Chinese Medicine into Bioactivity Space and Pathways Annotation Improves Mechanistic Understanding and Discovers Relationships between Therapeutic Action (Sub)classes. *Evidence-Based Complementary and Alternative Medicine* **2016**, 2016, 25 pages

Contents

Abstract	i
Acknowledgements	iii
CHAPTER 1: Introduction	
1.1 A brief history of traditional medicine	2
1.2 Natural products in drug discovery	2
1.3 Approaches to identify modes-of-action of natural products	4
1.4 Ligand-based <i>in silico</i> target prediction	6
1.4.1 The principle behind ligand-based <i>in silico</i> target prediction	
1.4.2 Applications of <i>in silico</i> target prediction	
1.4.3 Other methods of <i>in silico</i> target prediction	
1.5 Systems biology to elucidate modes-of-action	13
1.6 Limitations of <i>in silico</i> target prediction	16
1.7 Structure of this thesis	17
CHAPTER 2: Global Mapping of Traditional Chinese Medicine into Bioactivity Space and Pathways Annotation Improve Mechanistic Understanding and Discovers Relationships Between Therapeutic Action (Sub-)Classes	
2.1 Introduction	18
2.2 Materials and Methods	21
2.3 Results and Discussions	27
2.4 Conclusion	49
CHAPTER 3: Exploring the Chemical Space and Bioactivity Space of Traditional Medicine from China, India and Malaysia for Treating Cancers	
3.1 Introduction	50
3.2 Materials and Methods	53
3.3 Results and Discussions	61
3.4 Conclusion	80
CHAPTER 4: Evaluating Synergistic Pairwise Combinations of Shexiang Baoxin Pill (SBP) For Coronary Heart Disease From Network Topology	
4.1 Introduction	81
4.2 Materials and Methods	84
4.3 Results and Discussions	94
4.4 Conclusion	114
CHAPTER 5: Concluding Remarks	115
References	117
Appendices	131

Chapter 1:

Introduction

1.1 A brief history of traditional medicine

Humanity is dependent on natural products and has been so for millennia. These plants, animal parts and minerals are used both to treat many diseases (1, 2). Indeed, the knowledge of curative plants may be traced back at least 60,000 years where remains of medicinal plants such as opium poppies, ephedra, and cannabis were documented during archaeological excavations of Neanderthals burial sites at Shanidar in Iraq (3). The discovery of natural products with healing properties posed a massive challenge to early humans. Empirical discoveries of both healing and harmful properties of natural products resulted from trial and error, which can be fatal when experimenting poisonous natural products. Progressively, the knowledge of empirical discoveries was collected in the form of traditional medicine (4). The first written record of the use of natural products dates from 2600 BC of Mesopotamia, in which hundreds of clay tablets in cuneiform described, among others, the oils of *Cedrus* species (cedar) and *Cupressus sempervirens* (cypress), *Glycyrrhiza glabra* (licorice), *Commiphora* species (myrrh), and *Papaver somniferum* (poppy juice) were used for treating cold, cough, and inflammation (2, 4, 5). “Ebers Papyrus”, an ancient Egyptian pharmaceutical record (1500 BC), documented not only plants but also animal parts and minerals to constitute its 700 drugs (2, 4, 5). In other parts of the world, the ancient Greeks also documented the compilation of the uses of natural products for medicinal purposes (2, 4, 5). In Asia, traditional Chinese medicine (TCM) and Ayurveda (India) have been developed. These are two of the most extensively documented traditional medicines and so are considered in detail in this study (2, 4, 5). The practice of traditional medicine, over thousands of years, matured into systematic healthcare. Traditional medicine is defined by the World Health Organization (WHO) as sum total of the knowledge, skills, and practices based on the theories, beliefs, and experiences indigenous to different cultures, whether explicable or not, used in the maintenance of health as well as in the prevention, diagnosis, improvement or treatment of physical and mental illness (6). Ultimately, the goal of traditional medicine is to apply a holistic approach to maintain the overall balance of a human body.

1.2 Natural products in drug discovery

The identification of chemical compounds for observed therapeutic effects is still an ongoing process, despite the historical uses of natural products in traditional medicines for treating various diseases. The elucidation of chemical compounds for an observed therapeutic

effect was first reported in 1775 from the extract of *Digatalis purpurea* that was previously observed to treat arrhythmia (7). Subsequently, more compounds have been successfully discovered such as salicin, quinine and penicillin to name a few (5, 8). Some plant-derived drug have been developed with the same medicinal values of their sources (9). **Table 1.1** lists some representatives of drugs based on their ethnomedicinal uses. The previous historical success could be partly responsible for the results of a recent report, which indicated that at least 40% of the 1,562 drugs that were approved between 1981 to 2014, were natural products or inspired by natural compounds (10). Scientifically, natural products have been exploited in drug discovery, as they inherently offer a large structural diversity, much of which has yet to be explored (1). In addition, natural products offer a biological readiness to bind to many molecular targets resulting from the interactions with multi-enzymes during their synthesis, thus mimicking endogenous metabolites (1, 5). Natural products also show better solubility as, on average, they retain relatively low *log P* values, which in order to do so natural products, generally, have higher molecular weights to contain more hydrogen bond acceptors (1, 11). Given these points, natural products, with or without known indications from traditional medicines could provide a valuable alternate therapeutic space to explore new molecular entities leading to more efficacious drugs.

Table 1.1: A representative list of drugs that were developed based on the indications of their ethnomedicinal uses (9).

Drug	Action/Clinical use	Plant source	Traditional medicine
Digitoxin	Cardiotonic/arrhythmia	<i>Digatalis pupurea L.</i>	Europe
Asiaticoside	Wound healing/Tissue repair	<i>Centella asiatica</i>	Asia, Southeast Asia
Khellin	Asthma/Bronchodilator	<i>Amni visnaga</i>	Egypt
Scopolamine	Narcotic/Sedative	<i>Datura metel</i>	Arab
Theophylline	Diuretic/Diuretic	<i>Camellia sinensis</i>	Sri Lanka

Although natural products have contributed a fair share in drug discovery (10), they are still undervalued for various reasons. The first is technical: natural products are difficult to be fully synthesised due to their complex structures. This difficulty cannot be easily avoided by isolating compounds as the method of isolation could create a supply issue for target validation as well as being expensive in both time and finance (1, 12). Second, the shift from cell-based assay to target-based assay has discounted natural products from drug discovery process due to the incompatibility issue of natural product extracts in high throughput screening (HTS). The extracts contain a mixture of compounds thus, interfering with the process of elucidating mode-of-action (MOA) (12-14). Third, the rise of combinatorial chemistry, which allows the synthesis of large numbers of structurally distinct molecules, is

viewed to increase the chances of finding a ‘hit’ alongside to repress known issues posed by natural products (1, 12, 13). The limitations described above might also contribute to less attraction of looking into traditional medicine. In this study, we aim to propose the MOA of traditional medicines by exploring and assessing the applicability of *in silico* approaches towards this goal.

1.3 Approaches to elucidate mode-of-action

Initially, phenotypic-based screening was a primary focus in drug discovery, where compounds were tested on organisms, tissues or cells showing a desired phenotype (15, 16). The prior knowledge of ‘target’ is not required in this method because a compound is simply considered a ‘hit’ if it elicits a desired phenotype and a ‘miss’ if not (15, 16). With the advent of HTS technology and the Human Genome Project, the strategy in drug discovery shifted to target-based screening (16, 17). One of the strengths of target-based screening is that the target is readily identified and validated, therefore, explaining the MOA. This approach can expedite the drug discovery process but does so with a significant drawback that is to interpret the therapeutic effect of a compound, it must be considered in a relevant disease model (16, 17). Hence, phenotypic screening can be regarded to show higher potential to find a ‘hit’ but target identification is still required to understand the MOA. The target identification approaches can be described from the field of genetics, proteomics, and computational inference methods and reverse pharmacology (17-19). A brief description of each approach can be found in **Table 1.2**. In this study, we applied ligand-based *in silico* method to predict targets of traditional medicine compounds in order to rapidly improve mechanistic understanding of their ethnomedicinal uses.

Table 1.2: The four different approaches for target identification as described by Schenone et al. (17) , Chan et al. (18) and Vaidya A. (19)

Field of Approach	Description
Genetics	The approach identifies a target by deleting a gene to reduce the expression of protein that is hypothesised to be modulated for the desired phenotypic effect. An example of such approach is gene-knockout organism. However, the compound could also modulate a different target of the same pathway for the same desired phenotypic effect.
Proteomics	The approach identifies a target based on the physical interaction between a protein and a chemical compound. One of the classical approaches is the affinity chromatography where compounds of interest are immobilised on a column. The compounds are incubated with proteins extracts and non-specific bound proteins are moved by extensive washing. The bound proteins are eluted with excess free drug or under highly denaturing conditions and characterised by SDS-PAGE and mass spectrometry. Although, the method could provide the relative binding specificity, it is difficult to be implemented in HTS.
Computational inference methods	The approach is able to identify targets based on two methods; ligand-based and structure-based using computational algorithms. To put simply, the ligand-based method incorporates chemical structures to predict targets while the structure-based method relies on the three-dimensional structure of a protein to predict binding affinity. Whilst the method is fast, experimental analysis is still required to validate the predicted targets.
Reverse pharmacology	The approach identifies the target through genomic, proteomic and metabolomics studies. The potential hits are screened through target-binding assays where the highly selected hit the molecular target is identified. The highly selected hit undergoes <i>in vivo</i> studies significantly show the desired physiological effect.

1.4 Ligand-based *in silico* target prediction

1.4.1 The principle behind ligand-based *in silico* target prediction

Rapid development in both combinatorial chemistry and HTS has allowed the explosion in number and size of compound libraries and bioactivity databases(20-22). These two types of databases provide an interface to study both biology and chemistry concurrently using informatics approaches, which has given rise to computational chemogenomic (23, 24). Some of the examples of the compound libraries and/or bioactivity data can be found in **Table 1.3** and **Table 1.4**. In addition to aforementioned databases, compound libraries of natural products have also kept arising (**Table 1.5**).

Table 1.3: The databases that stores compounds and their structural information.

Database	Description	Size of database (as of Jan 2017)	Ref.
ChemSpider	The database is publicly accessible database that contains chemical structures and their physical and chemical properties, spectral data, synthetic methods, known reactions, safety information and systematic nomenclature. (http://www.chemspider.com)	58 million structures	(25)
ZINC	The database is a collection of chemical compounds. (http://zinc15.docking.org)	The current version 15 contains more than 120 million compounds.	(26)

Table 1.4: The databases that contain compounds and their structural information together with their bioactivity information.

Database	Description	Size of database (as of Jan 2017)	Ref.
Binding DB	The database is publicly accessible that stores experimental protein-small molecule interaction data. The data is curated from scientific articles and US patents. (www.bindingdb.org)	A total of 1,325,129 bioactivity data (US patents, journals, other databases) of 6,914 protein targets and 589,218 small molecules.	(27)
ChEMBL	A public drug discovery database consisting of small bioactivity molecules derived from scientific journals. (https://www.ebi.ac.uk/chembl/db)	The current version 22 consists of 14,371,197 bioactivity data (HTS) of 1,686,695 compounds and 11,224 targets.	(28)
PubChem	The public database contains chemical compounds and their activities against different biological assays. (https://pubchem.ncbi.nlm.nih.gov/#)	A total of 232,660,585 bioactivity data (HTS) of 10,341 protein targets and 94,475,253 compounds.	(29)

Table 1.5: Compound libraries of natural products that contain their structural information, physicochemical properties or bioactivity activities.

Database	Description	Size of database (as of Jan 2017)	Ref.
Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP)	The database is publicly accessible and contains chemical compounds that link to their herbs and their associated targets and diseases. Associated targets and diseases are retrieved from DrugBank (30) and TTD database (31) and PharmGKB (https://www.pharmgkb.org/). (http://ibts.hkbu.edu.hk/LSP/tcmsp.php)	29,384 structures from 499 Chinese herbs registered in the Chinese pharmacopoeia and they are associated to 3,311 targets and 837 diseases.	(32)
Naturally occurring Plant based Anti-cancerous Compound-Activity-Target Database (NPACT)	The database is publicly accessible and each record provides information on their structure, properties (physical, elemental and topological), cancer type, cell lines, inhibitory values (IC ₅₀ , ED ₅₀ , EC ₅₀ , GI ₅₀), molecular targets, commercial suppliers and drug likeness of compounds. (http://crdd.osdd.net/raghava/npact/index.html)	1,574 compounds curated from 762 references, which include 353 cell lines of 19 cancer types.	(33)
Super Natural II	It is publicly available database of natural compounds with additional information such as MOA, pathways, and toxicity. (http://bioinf-applied.charite.de/supernatural_new/index.php?site=home)	Approximately 326,000 unique compounds, collected from 16 suppliers.	(34)

Ligand-based *in silico* target prediction (“*in silico* target prediction” will be used hereinafter) applies chemogenomics that aims to systematically study the relationship of compounds and bioactivity space by annotating compounds to their targets and implementing algorithms to measure compound-target classification, which is used to predict targets of an orphan compound (24, 35-37) (**Figure 1.1**). Two basic assumptions are applied in chemogenomics (38):

1. Structurally similar compounds modulate similar targets
2. Similar targets share similar patterns (binding sites)

From these assumptions, chemogenomics is able to identify molecular recognition between all possible ligands and all proteins in a biological system (23, 39). There are two main elements to *in silico* target prediction:

1. Generating chemical compound descriptors
2. Developing a machine learning model to identify patterns in the compound descriptors

The first element characterises a chemical compound using properties called descriptors (23). Descriptors can be described through their dimensionality from one dimensional (1D) through to three dimensional (3D) (36, 40, 41). 1D descriptors describe the physicochemical properties of a molecular structure such as molecular weight, polarity, and molar refractivity (36, 40, 41). The 1D descriptors are easy to interpret but using only one type of 1D descriptor to characterise set of compounds does not offer sufficient discriminating power (36, 40, 41). Therefore, it is usually used in combinations with other 1D descriptors (36, 40, 41). 2D descriptors encode molecular connectivity, which can be described as topological indices and fingerprints (36, 40, 41). The topological indices are integers calculated from the 2D-graph representation of a compound (36, 40, 41). Fingerprints are encoded as bit strings, which the descriptors predefine the presence (1) or absence (0) of fragments or substructures in a particular chemical structure (36, 40, 41). 3D descriptors describe the shape of the chemical compound such as pharmacophore-type representation (36, 40). The 3D descriptors may be able to provide information on the binding properties of the compounds (36, 40). However, calculating 3D descriptors could require large computational time for a large set of compounds (36, 40). The descriptors are used to identify similarities between molecules. The similarity measure between two compounds, using a pair of bitstring, is usually computed using Tanimoto coefficient (Tc) or Jaccard Index. The Tc can be explained from the equation below:

$$Tc = \frac{c}{a + b - c}$$

(Equation 1.1)

a is number of features present in compound A, b is number of features present in compound B and c is number of features present both in compound A and B. The Tc value is between 0 and 1 with value closer to 1 shows higher similarity between the two compounds. Altogether, the first part describes the ligand space.

The second element is mapping the features of the first component to an output property. This is done by building a chemogenomic database or ligand-bioactivity space from selected bioactivity databases (42). The bioactivity data is divided into ‘training set’ and ‘test set’ (42). In general, the ‘training set’ contains experimental results of x compounds acting on y targets, which are then ‘trained’ using an appropriate algorithm (23). The algorithm, such as

Naïve Bayes classifier, Random Forest and Support Vector Machine, will represent the training set in a form of classification model, for the machine to learn, which eventually enables target prediction of novel a compound (23, 36). The details of the algorithms will be explained in Chapter 2 (Naïve Bayes classifier) and Chapter 3 and 4 (Random Forest) as two different versions of *in silico* target prediction were used. Targets are predicted based on the similarity of the compounds in the training set to modulate the same targets (23). The performance of the model is validated by correctly predicting the targets of a ‘test set’, which is also known as internal set validation or the targets of a different bioactivity database, which is also known as external set validation (42). Therefore, by manipulating a large bioactivity data, *in silico* target prediction is perceived as one of the possible methods to potentially predict targets of TM compounds to eventually suggest their MOAs. The applications of *in silico* target prediction have been reported in the literature using different chemical fingerprint, different training sets of data and different classification algorithms, which will be further discussed next.

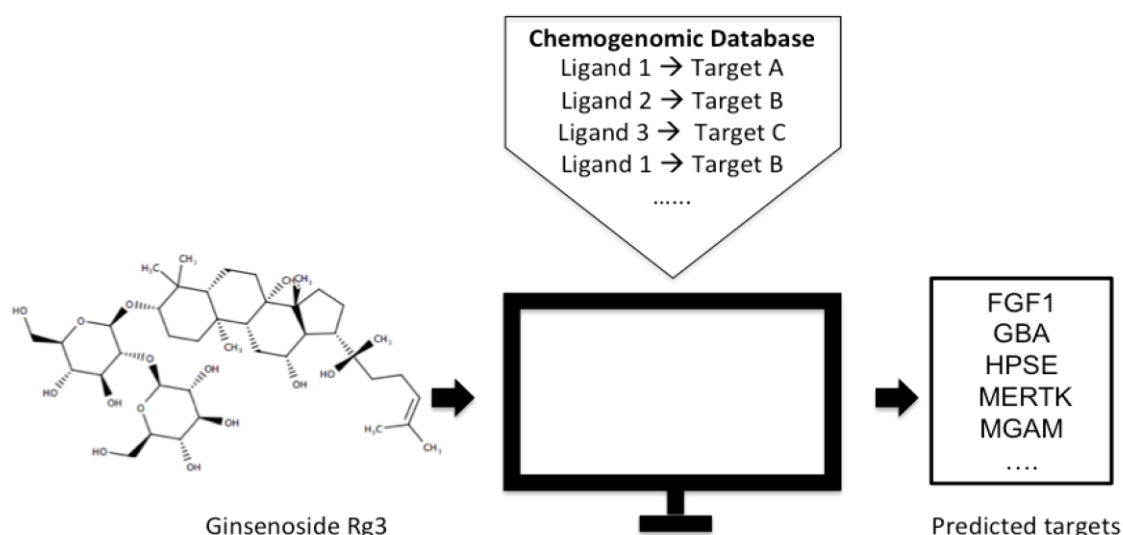


Figure 1.1: Workflow of *in silico* target prediction tool using chemogenomic approach. A set of bioactivity compounds is compiled to build the chemogenomic database. An appropriate algorithm is used to learn the classification of the chemogenomic database. Then, using a novel compound as enquiry, the target(s) of a novel compound is predicted by comparing the similarity of the novel compound to the compounds in the chemogenomic database to produce a list of putative targets.

1.4.2 Applications of *in silico* target prediction

One of the earliest studies published on target prediction is the Prediction of Bioactivity Spectra for Substances (PASS) method (43). The model was developed using

structure-activity relationships for over 300,000 biologically active organic compounds representing for more than 4,000 types of biological activity collected from databases and literature. The structures were characterised by a set of descriptors of multilevel neighbourhood atom (MNA) and the prediction was based on Bayes estimates. The model's performance showed it achieved average accuracy above 95% in the leave-one-out cross validation approach.

Nidhi et al. used a large chemogenomic dataset to develop the *in silico* target prediction (44). In this work, ligand-target pairs from the World of Molecular BioAcTivity (WOMBAT) database were extracted and the compounds were encoded using Extended-Connectivity Fingerprint (ECFP). A multiple-category Laplacian naïve-Bayes algorithm was trained on the model consisted of 964 target classes. The model was employed to predict the top three most likely targets for all compounds from a different database, the MDL Drug Database Report (MDDR). The results showed that, on average for ten MDDR activity classes with known targets, the top three targets were predicted correctly for 77% of the compounds.

In the Similarity Ensemble Approach (SEA), the target prediction was developed by grouping and relating the protein targets based on chemical similarity among their ligands, which were quantitatively measured using an algorithm adapted from BLAST (45). The model contained over 65,000 ligands and 246 targets of MDDR. The study discovered that the chemical structure of methadone, loparamide and emitidine showed close structural similarity to ligands for the muscarinic M3, α 2 adrenergic and neurokinin NK2 receptors as their off-targets. The experimental validations indicated the binding affinities of the drugs to antagonise the targets were in micro molar ranges.

In contrast to different types of bioactivity classes, Mestres et al. conducted a study that focused on one class of target protein, nuclear human receptor, (NHR) (46). The model consisted of 2,033 molecules and 25 nuclear receptors extracted from literature. The molecules were characterised using Shannon Entropy descriptors (SHED), which described the topology of molecules by mapping the distributions of atom-cantered feature pairs. A nuclear receptor network derived from a matrix of minimum Euclidean distance of SHED profiles. The profiling of the nuclear receptors was tested against four external focused libraries of proteases, kinases, ion channel and GPCR, and 2,944 NHR drugs. The results showed that between 4.4% and 9.7% of the focused libraries and 5.2% of the drugs having affinities to binding NHR. Thus, this method does not only identify putative targets but also potential off-targets.

In regard to traditional medicine, the application of *in silico* target prediction was reported for TCM and Ayurveda compounds. Mohd Fauzi et al. employed a predictive model that was generated from ChEMBL bioactivity database version 10.0, which consisted of 155,208 molecules across 894 targets (47). A Naïve Bayes classifier was utilised to train the model and the compounds were characterised based on ECFP_4 (48) with a diameter of four bonds. The model was validated using five-fold cross validation and achieved a recall of 68%. TCM compounds of ‘tonifying and replenishing’ medicinal class were successfully connected to their relevant targets that were associated to hypoglycaemic effect and Ayurveda cancer compounds were predicted to modulate primary anti-cancer targets. Both predictions were validated by literature. Given the previous studies, *in silico* target prediction can be extended to predict targets of traditional medicine by linking to their ethnomedicinal uses (**Figure 1.2**).

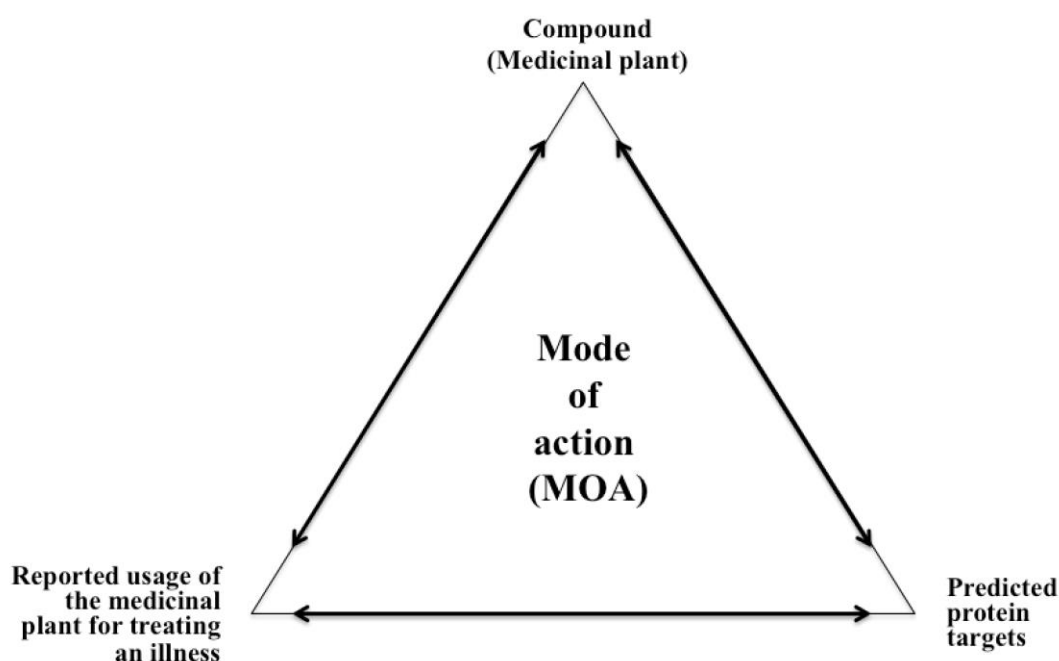


Figure 1.2: Visualization of the link between a compound, target, and an indication. The compound’s mode-of-action can be elucidated from the targets predicted using an *in silico* target prediction method, which can be rationalised by the reported usage of the medicinal plant for treating an illness. The three elements are strongly interconnected with each other.

1.4.3 Other methods of *in silico* target prediction

In silico target prediction based on machine learning methods are not the only available tool. Various methods are available for *in silico* target prediction such as protein/ligand panel docking, chemical similarity searching and the analysis of “bioactivity spectra” (49). Protein panel docking is one of the earliest tools that has been widely used in traditional medicine, where a compound is docked to a wide panel of potential proteins and the proteins are

subsequently ranked based on calculated binding affinity scores (50). Chen et al., used INVDOCK, a ligand binding inverse docking strategy, against nine TCM compounds, genistein, ginsenoside Rg1, quercetin, acronycine, baicalin, emodin, allicin, catechin and camptothecin (50). The study discovered that over 50% of the predicted targets had relevant literature support to suggest the compounds' MOAs. In a different study, where the *in silico* target prediction was further validated experimentally, Zhang et al. identified putative targets of 19 compounds isolated from two medicinal plants in TCM, which were used for the treatment of diabetes and inflammation using a reverse docking approach, TarFisDock (51). The natural products showed moderate inhibitory activities against the most frequent target candidate, dipeptidyl peptidase IV (DPP-IV) with IC₅₀ values ranging from 14.14μM to 113.76μM in an *in vitro* enzyme assay (51). Although this protein panel docking method requires only the chemical structures of the putative active ingredients, it is limited to high quality protein structures and the accuracy of the docking programs used (52).

Chemical similarity searching is based on the “Molecular Similarity Principle” (41, 49). In this method, the query compound is compared to a database of compounds with known targets and the compounds are represented as chemical fingerprints which the similarity between the two compounds is measured (52). In a study by Cleves et al., (53), out of 1000 drugs, 700 drugs modulate approximately 85 biological targets from three different analysis; drug and target similarities, ligand-based virtual screening and selectivity of ligand-based models. When the results were compared, the “chemical similarity” principle performed well when drug pairs sharing a target had higher similarity. Although this method does not require a training set, it does not prioritise which target is important for the query compound and prior knowledge of target class information is ignored (52).

A method based on bioactivity spectra uses information from the response to a compound across a series of biological readouts such as cell lines, targets or genes expression profiles, which creates biological activity spectra (49). As bioactivity spectra are related to the compound's structure, it can also be viewed as molecular descriptors and used for target prediction (52). For example, Chen et al., developed bioactivity profile similarity search (BASS) using 4,296 compounds that were tested against the NCI-60 cell lines (54). The annotated targets for all the compounds were extracted from public databases. Potential targets of a query compound were calculated based on bioactivity profile similarity between the query compound and all reference compounds in the database. The model demonstrated 44.8% successful predictions from 237 query compounds with known target annotations. Although, this method could select a set of compounds with relevant bioactivity data, it requires the initial collection of experimental data across many targets, which can be difficult

to obtain (52). Given these points, ligand-based *in silico* target prediction could at least overcome some of these limitations by having a larger bioactivity space and statistically prioritised predicted targets.

1.5 Systems biology to elucidate mode-of-action

While target identification continues to be important to elucidate compounds' MOA, the targets by themselves do not represent the complete mechanism of the MOA. Biologically, a measure of a compound's response comes from systems biology (cells, tissues, animals) suggesting the target is not acting alone but interacting with neighbouring proteins (55, 56). The chain of interactions forms a dynamic biological network to produce phenotypic response (55, 56). The chain of interactions, which can be defined as a pathway, is the collection of proteins acting in collaboration that is part of a particular function or process (57). Multiple pathways of shared targets and proteins can be embedded into a biological network to provide a more comprehensive view of how systems biology functions to better understand the compound's MOA (55, 56). However, analysing pathway in isolation from the network has also been done, in some cases, as pathway annotations are able to describe the compound's MOA (57). The motive to include systems biology in drug discovery derives from a growing understanding of the physiology of a disease (58). The manifestation of a disease involves multiple interactions between targets of various pathways in a biological network (58). Therefore, by capturing beyond target space, the strategy is seen to be able to elucidate the MOA of a compound from the pathogenesis of a disease. In comparison, this concept aligns closely to the holistic approach of traditional medicine that aims to restore a healthy state of a patient with localising symptoms rather than to treat the source of the illness.

Efforts have been seen in *in silico* drug discovery to include systems biology. One of the continuous efforts is the emergence of public databases of molecular interactions. At target level, several databases of interacting proteins have been compiled. The protein-protein interactions (PPI) databases are regularly collected from literature curation and high-throughput protein interaction assays and in some cases, the databases include predicted interactions as well (59, 60). These PPIs are also put into biological contexts by creating pathway databases, which can have different emphasis and coverage (57, 60). **Table 1.6** and **Table 1.7** list the statistics/properties of the publicly available PPI and pathway databases. Consequently, the information stored in both types of databases can be combined to understand how the organisation at the molecular level contributes to cellular and organism phenotypes.

Table 1.6: The statistics of protein interactions (*Homo sapiens* only) databases.

Acronym	Database full name	Proteins (as of Jan 2017)	Interactions (as of Jan 2017)	Publications (as of Jan 2017)	Ref.
IntAct	IntAct Molecular Interaction Database	46,895	208,595	7,182	(61)
HPRD	Human Protein Reference Database	30,047	41,327	453,521	(62)
BioGRID	Biological General Repository for Interaction Datasets	20,931	278,358	25,759	(63)
MINT	Molecular Interaction Database	10,807	61,143	8,644	(64)

Table 1.7: The properties of pathway databases.

Acronym	Database full name	Coverage	Interactions (as of Jan 2017)	Ref.
KEGG Pathway	Kyoto Encyclopaedia Gene and Genome Pathway	Represents collections of manually drawn pathway map of metabolism, genetic information processing, cellular processes, organismal systems, human disease, and drug development. (http://www.genome.jp/kegg/pathway.html)	481,160	(65)
Signalink2.0	Signalling pathway resource with multi-layered regulatory networks	Includes signalling pathway cross-talks, transcription factors, miRNAs and regulatory enzymes. (http://signalink.org/)	275,359 (only for <i>Homo sapiens</i>)	(66)
Reactome	Reactome Pathway Database	Includes core human pathways such as DNA replication, transcription, translation, cell cycle, metabolism, and signalling cascades. (http://www.reactome.org/)	209,988	(67)

A few examples are discussed here to illustrate the inclusion of pathways to understand compound's MOA. For instance, a study by Liggi et al., demonstrated that enriched KEGG pathways such as “axon guidance” and “endocytosis” of predicted targets could be linked to the frequently observed phenotype, pigmentation, of *Xenopus laevis* that were previously treated with the NCI Diversity Set II compounds (68). The functions of the targets and pathways to pigmentation were validated by literature (68). Relating to traditional medicine, Liu et al., generated a compound-target network of nine compounds that were identified to be

abundance in Reduning Injection TCM formulation used in combatting inflammation (69). Inflammation-related pathways were first extracted from literature, which also identified a few of major drug targets. The compound was connected to the drug targets if they were predicted targets of the compounds. Related targets, such as mitogen-activated protein kinases (MAPK1 and MAPK14), were identified from the network topology. Experimental results showed that the TCM formulation modulated the MAPK pathway to exert its anti-inflammatory effects. Therefore, by annotating pathways of the targets, the MOA of compounds could be understood better for the reported indications.

Because systems biology is more than discrete PPIs and pathways, thus, creating a biological network using either of both components could provide a more comprehensive insight of the compound's MOA. Depending on a research question and available data, a biological network is regularly modelled using two different approaches (70).

1. Top-down or data-base approach
2. Bottom-up or knowledge-base approach

In a top-down approach, experimental data is extracted to infer a biological network. Various methods can be employed to build the network namely statistical (correlation and regression) and probabilistic methods (Bayesian, logic models, and ordinary differential equations), in which the gene expression data can be either static or time-course (70). As a result, the approach attempts to correlate the molecule's behaviour and underlying interactions in the genome-wide studies (71). For instance, Liu et al., applied correlation statistical method by generating a network using weighted gene co-expression network analysis (WGCNA) from gene expression samples of breast cancer patients who received Tamoxifen treatment. The aim was to identify the biomarkers in order to understand the underlying mechanism for Tamoxifen resistance (72). The analysis used unsupervised clustering to identify functional modules and significant genes, which were validated using independent samples of breast cancer survivors treated with Tamoxifen (72). The results indicated five hub genes (CDK1, DLGAP5, MELK, NUSAP1, and RRM2) were strongly related to poor survival (72). Therefore, the top-down approach could facilitate the causal relationship between phenotype and a drug.

Contrary to the top-down approach, the bottom-up approach starts with single components or the proteins of interest as the building blocks of the network in order to simulate the dynamic properties of the resulting system (70). The physical interactions between these proteins could be retrieved from PPI or pathway databases, which eventually the interactions are combined to build a biological network (70). The outcome of this method maps all possible connections between proteins in the network because the overlapping

information between the databases is usually small (59, 60). To illustrate this approach, Zhu et al., used differentially expressed genes measured between non-failing heart and ischemic dilated cardiomyopathy (ICM) patients, in addition to validated cardiac myocyte proteins to build a representative network of dilated cardiomyopathy (73). The PPI were retrieved from DIP, BIND, Prolinks, KEGG and HPRD (73). The network was divided into four layers; extracellular, plasma membrane, cytoplasm and nucleus, which Gene Ontology (GO) analysis showed the over-represented biological processes were in the cytoplasm and nucleus layer (73). Network analysis revealed that the Janus family tyrosine kinase-signal transducer and activator of transcription (Jak-STAT) signalling pathway might play an important role in the development of ICM (73). Subsequently, treating related drugs in *in vitro* experiments could validate the biological processes discovered. Together, the applications discussed of systems biology are observed to prospectively fill the gap between chemistry and biology, thus potentially helping the drug discovery process in the future.

1.6 Limitations of *in silico* mode-of-action methods

Despite the applications of *in silico* target prediction able to elucidate the MOA, there are several limitations. First, the “Molecular Similarity Principle” is not always true. There are cases where molecules that are similar in structure but exhibit very different activities, which is known as the “activity cliff” (74). For example, a change in one functional group might prevent a compound from binding to the same target. To confirm the predicted target is modulated by the compound, a molecular docking method is proposed in order to learn about the molecular interactions between the compound and predicted target. Second, the data in the training set contains inconsistent results. The inconsistent results may be due to different experimental design and interpretation, such as insufficient description of assays, making it difficult to compare two similar assays from different experiments (75). This can be minimised by selecting bioactivity data based on several confidence criteria. Third, the predictions are still limited to the target space in the training set. Hence the model usually covers a fraction of proteins in the human genome. However, as more bioactivity data becomes available, the size of target space can be further expanded. Fourth, the algorithm is unable to report whether a novel compound either activates or inhibits the predicted targets, as this can be experimentally influenced by ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties (68). Therefore, only a general justification could be established between the predicted targets/pathways and the phenotypic response. Finally, relating to the PPI and pathway databases, despite the large size of most databases, their curation is error-prone and the protein interactions coverage is still incomplete as many parts

of the systems biology have yet to be studied (57). Hence, these limitations remain to potentially affect the analysis of *in silico* MOA's results.

1.7 Structure of this thesis

Given that, traditional medicines are known for the observed efficacies, the applicability of *in silico* target prediction in elucidating the MOA of the compounds, of one or different types of traditional medicine, is explored. The work is presented in self-contained chapters. Chapter 2 discusses the mechanistic relationship between 46 TCM therapeutic action subclasses that each subclass describes a distinct healing property. In the subsequent chapter, Chapter 3, the work compares both chemical space and bioactivity space of three different traditional medicines, namely TCM, Ayurveda, and Malay traditional medicine (Malay TM) for treating cancers. The final work, Chapter 4, evaluates synergy of two compound combinations from a TCM formulation, Shexiang Baixin Pill (SBP), by mapping the compounds predicted targets onto a representative biological network. The formulation is widely known for treating coronary heart disease. Finally, Chapter 5 discusses the results and conclusions from previous chapters.

Chapter 2:

Global Mapping of Traditional Chinese Medicine onto Bioactivity Space and Pathway Annotation Improves Mechanistic Understanding and Discovers Relationships Between Therapeutic Action (Sub-)Classes

2.1 Introduction

Traditional Chinese medicine (TCM) has been practiced for thousands of years for the prevention and treatment of diseases using a unique system of theory, diagnosis and treatment (76, 77). The philosophical background of TCM is based on Yin and Yang, as well as the Five Elements (agents) theories (**Figure 2.1**). The Yin and Yang are the harmony of two opposite energies and the Five Elements describe the five interdependent functional organs; heart, liver, spleen, lung and kidney with its own Yin and Yang (78, 79). When a human body suffers from a disease, the dynamic balance and the relationship of the five elements are disturbed. Hence, to rectify the disturbance, TCM applies a holistic approach with the key therapeutic principles being “*Zheng*” (meaning syndrome or pathological patterns seen in patients) and “*Fufang*” or “*Fang Ji*”, (meaning compound formulations consisting of *materia medica*) (80-86). Note that a TCM *Fufang* are primarily based on medicinal plants, but may also contain fungi (e.g. *Ganoderma lucidum*), mineral (realgar), and occasionally animal products (e.g. *Calculus bovis*). Chinese medicines can be organised into several classifications such as therapeutic actions, source of the medicine and internal organs (87). In this study, the classification of the Chinese medicines follows the therapeutic actions, of which some also possess sub-classifications based on clinical applications recorded by TCM monographs (88). A combination of two or more Chinese medicine categories make up a treatment formulation, which then contains a considerable number of chemical compounds (89).

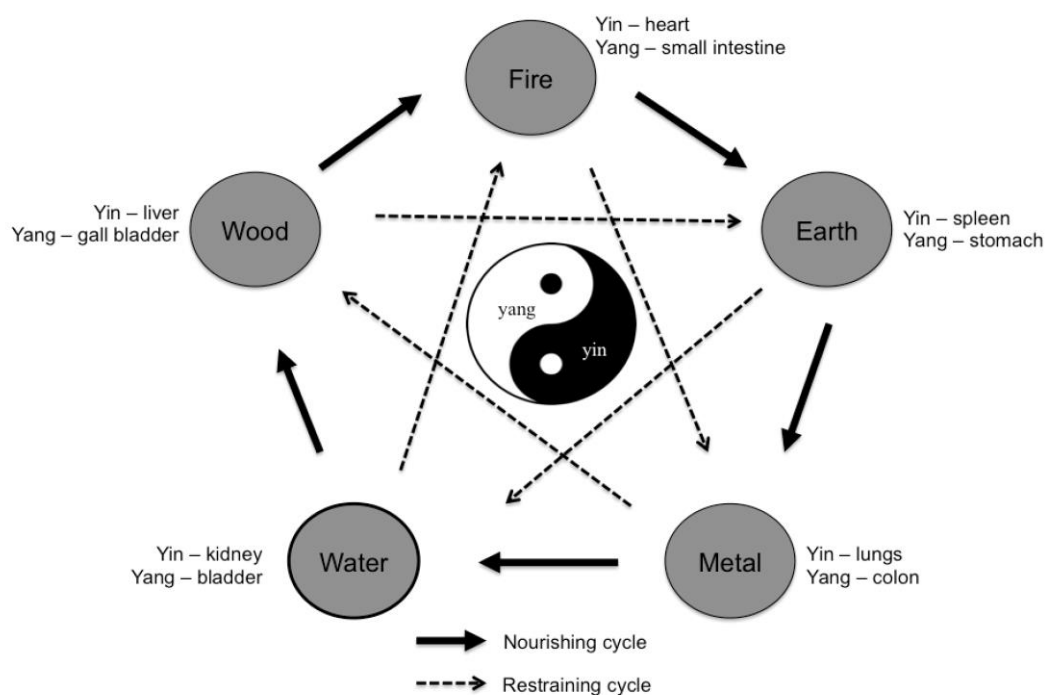


Figure 2.1: The philosophy of TCM defines yin and yang as opposite energies that complement the natural world. Water, earth, metal, wood and fire are the five elements that organise all natural phenomena into five patterns. Each element has its own organs of yin and yang characteristics. Each element can either nourish or restrain one another to keep a dynamic balance among the elements. TCM compounds of *materia medica* are used to correct the imbalance of yin and yang in the human body.

The mixture of compounds in the formula works through the therapeutic principle *Jun-Chen-Zuo-Shi*, by maximizing the therapeutic effects and minimizing the side effects (76, 90, 91). Based on the healing/pharmacological properties and constituents of each medicine, the Jun (emperor) component is the principal compound in the mixture targeting the major symptom of the disease. There are only a few varieties of Jun medicines that are administered as a single formula, usually in large doses. The Chen (minister) components synergise with Jun to strengthen its therapeutic effects, and may also treat secondary symptoms. The Zuo (assistant) medicine reduces or eliminates possible adverse or toxic effects of the Jun and/or Chen components, while also enhancing their effects and sometimes treating secondary symptoms. Finally, the Shi (courier) components facilitate delivery of the principal components to the lesion sites, or facilitate the overall action of the other components (92, 93). Therefore, at the molecular level, a TCM formula, which is a multi-component and multi-target agent, is assumed to modulate a series of protein targets in an integrative manner to harmonise the body system (94). In brief, TCM is a well-structured system from diagnosis to healing, whose theories and medicines are rationally connected and interdependent.

TCM contains a plethora of compounds, which originate from the biodiversity of natural product medicines, which is a rich resource for discovering new TCM-based drugs.

However, to develop a novel TCM-based drug still remains challenging. One of the factors that contribute to making it a challenge is an undefined medicine concoction. The characterisation of the complex formulation by using methods to isolate the compounds is an exhaustive task that is very time consuming (95, 96). Although many compounds have been isolated from Chinese herbs (97, 98), their MOAs, in many cases, are still not yet understood at the molecular level (96). Another challenging issue in TCM is measuring the efficacy, pharmacokinetic-pharmacodynamic profiles and dose-efficacy relationship of multiple compounds simultaneously, both *in vitro* and *in vivo* (95, 96, 99). However, as shown in this work, particularly related to MOAs, it is now possible to suggest the MOA of TCM compounds using *in silico* target prediction. For example, Zehler et al, applied inverse screening method for three indirubin (TCM active principle in against chronic myelogenous leukaemia) derivatives, 5-bromo-indirubin-3'oxime (5BIO), 6-bromo-indirubin-3'oxime (6BIO) and 7-bromo-indirubin- 3'oxime (7BIO) in order to identify their potential kinase targets (100). The predicted target phosphoinositide-dependent kinase 1 (PDK1) was functionally supported by an *in vitro* kinase assay, which showed inhibition activity of 6BIO at 1.5 μ M (100).

The motivation of the current study is prompted by the success to link TCM and Ayurveda compounds to the predicted targets that were relevant for the indications of various therapeutic action classes in TCM and cancer in Ayurveda (47). The aim of the study was to extend the analysis of the targets prediction compounds of TCM therapeutic action classes including their subclasses by first generating a hierarchical clustering. The first part of the study was to understand the MOAs of the subclasses from the predicted targets as well as from annotated KEGG pathways because a target alone is insufficient to provide a full biological profile towards the effect of the ligand on a biological system (65, 101). In the second part, bioactivity spaces of all therapeutic action (sub-)classes were compared in order to understand relationships between the clusters. Hence, the global mapping of TCM compounds explored in this work, based on the therapeutic action (sub-)classes, does not only provide the better insight of the MOAs of the TCM compounds but also describes for the first time the relationship between the therapeutic action (sub-)classes.

2.2 Materials and methods

2.2.1 Dataset and dataset preparation

TCM compounds were obtained from TCM Database@Taiwan (88) in structure-data (SD) format. A total of 13,091 compounds from 46 different therapeutic action (sub-)classes were imported into MOE (102). To prepare structures for further analysis, covalently bound Group I metals were disconnected into ionic representation, while keeping only the largest molecular fragments, neutralising the compound by removing salts and adjusting the hydrogens and partial charges using MMFF94 (modified) partial charges. The duplicates from each therapeutic action (sub-)class were removed, resulting in a total number of 10,749 distinct compounds. The list of classes with the final number of compounds for each (sub-)class is provided in **Table 2.1**.

Table 2.1: The list of therapeutic action classes, subclasses and their number of respective Chinese medicine and compounds. A total 10,749 compounds from 46 therapeutic action subclasses were included in the analysis presented in this work.

Chinese medicine Class	Chinese medicine Subclass	Chinese medicine (sub-)class (Chinese Names)	Abb.	NoH	NoC
Exterior releasing	Wind cold dispersing	Sàn hán jiě biǎo yào (散寒解表药)	ER - wind cold	21	538
	Wind heat dispersing	Qīng rè jiě biǎo yào (清热解表药)	ER -wind heat	22	413
Heat clearing medicinal	Heat clearing and blood cooling	Liáng xuè huó xuè yào (凉血活血药)	HC - blood cool	14	99
	Heat clearing and dampness drying	Qīng rè lì shī yào (清热利湿药)	HC - damp	11	264
	Deficiency	Qīng xū rè yào (清虚热药)	HC - def	10	186
	Heat clearing and detoxicating	Qīng rè jiě dú yào (清热解毒药)	HC - detox	54	1029
	Heat clearing and fire purging	Qīng rè xiè huǒ yào (清热泻火药)	HC - fire purge	18	234
Purgative medicinal	Laxative medicinal	Rùn xià yào (润下药)	Purg - lax	3	27
	Offensive purgative	Gōng xià yào (攻下药)	Purg - off	6	54
	Drastic (purgative) water-expelling	Jùn xià zhú shuǐ yào (峻下逐水药)	Purg - water expel	14	206
Dampness resolving	Water draining and anti-icteric	Lì shī tuì huáng yào (利湿退黄药)	Damp - antiicteric	6	189
	Water draining and strangury resolving	Lì niào tōng lín yào (利尿通淋药)	Damp - stran	15	133
	Water draining and swelling dispersing	Lì shuǐ xiāo zhǒng yào (利水消肿药)	Damp - swell	13	265
Qi regulating		Lǐ qì yào (理气药)	Qi	36	699
Digestant medicinal		Xiāo shí yào (消食药)	Digest	8	146
Hemostatic medicinal	Astringent hemostatic	Shōu liǎn zhǐ xiě yào (收敛止血药)	Hemo - astringent	5	92
	Blood cooling	Liáng xuè zhǐ xiě yào	Hemo -	13	198

	hemostatic	(凉血止血药)	blood cool		
	Meridian warming hemostatic	Wēn jīng zhǐ xuè yào (温经止血药)	Hemo - meridian	2	146
	Stasis-resolving hemostatic	Huà yū zhǐ xiě yào (化瘀止血药)	Hemo - stasis	6	245
Blood activating and stasis resolving	Blood activating analgesic	Huó xuè zhǐ tòng yào (活血止痛药)	BASR - analgesic	7	487
	Blood breaking mass eliminating	Pò xiě xiāo zhēng yào (破血消癥药)	BASR - break	9	177
	Blood activating menstruation resolving	Huó xuè tiáo jīng yào (活血调经药)	BASR - menstrual	15	457
	Blood activating trauma curing	Huó xuè liáo shāng yào (活血疗伤药)	BASR - trauma	12	261
Cough suppressing and panting-calming	Clearing and Heat phlegm resolving	Qīng huà rè tán yào (清热痰药)	CSPC - heat	30	237
	Cold phlegm resolving and warming	Wēn huà hán tán yào (温化寒痰药)	CSPC - cold	19	147
	Cough suppressing and panting calming	Zhǐ ké píng chuǎn yào (止咳平喘药)	CSPC - panting	16	334
Tranquilizing	Heat nourishing tranquilizing	Yǎng xīn ān shén yào (养心安神药)	Tranquil - heat	1	145
	Settling tranquilizing	Zhòng zhèn ān shén yào (重镇安神药)	Tranquil - settle	6	1
Orifice opening		Kāi qiào yào (开窍药)	Orifice	7	68
Liver-pacifying and wind extinguishing	Extinguish wind to resolve convulsion	Xī fēng zhǐ jìng yào (息风止痉药)	LPWE - convulsion	8	85
	Liver yang calming	Píng yì gān yáng yào (平抑肝阳药)	LPWE - liver	7	22
Tonifying and replenishing	Blood tonifying	Bǔ xuè yào (补血药)	TR - blood	7	388
	Qi tonifying	Bǔ qì yào (补气药)	TR - qi	15	474
	Yang tonifying	Bǔ yáng yào (补阳药)	TR - yang	23	559
	Yin tonifying	Bǔ yīn yào (补阴药)	TR - yin	17	259
Astringent	Anhidrotic	Gù biǎo zhǐ hàn yào (固表止汗药)	Ast - anhidro	3	17
	Lung-intestine astringent	Liǎn fèi sè cháng yào (敛肺涩肠药)	Ast - lung	8	145
	Secure essence, reduce urination, and check vaginal discharge	Gù jīng suō niào zhǐ dài yào (固精缩尿止带药)	Ast - secure	6	125
Wind-dampness dispelling	Bone(sinew) strengthening	Qū fēng shī qiáng jīn gǔ yào (祛风湿强筋骨药)	WD - bone	5	44
	Heat clearing	Qū fēng shī rè yào (祛风湿热药)	WD - heat	8	175
	Cold dispersing	Qū fēng hán shī yào (祛风寒湿药)	WD - cold	13	309
Interior warming		Wēn lǐ yào (温里药)	Warm	13	457
Worm expelling medicinal		Qū chóng yào (驱虫药)	Worm	9	93
Emetic medicinal		Yǒng tǔ yào (涌吐药)	Emetic	3	9

Parasite destroying, dampness eliminating and itchiness relieving	Gōng dú shā chóng zhǐ yǎng yào (攻毒杀虫止痒药)	Parasite	8	81
Anti-malarial medicinal	Kàng nüè yào (抗疟药)	Malarial	4	30
Total compounds			10,749	

Abb. = Abbreviation

NoH = Number of Chinese medicine

NoC = Number of compounds

2.2.2 Target prediction

The processed molecular data was then subjected to a target prediction that was modelled using the Laplacian-modified Naive Bayes classifier, which detail of the target prediction algorithm can be found in (103). Briefly, this model contained 189,147 ligand-protein pairs extracted from ChEMBL v.14.0 (28) across 477 human targets which was used as the training set. The training set contained active compounds with reported activities ($K_i/K_d/IC_{50}/EC_{50}$) of at least 10 μ M with a confidence score of 8 or 9 and at least 20 compounds were available to associate the chemical features with a target class.

The molecular descriptors of the compounds were represented by Molprint2D circular fingerprints (104). In Molprint2D circular fingerprints (**Figure 2.2**), the atom environments of a molecule are directly calculated from molecular connectivity table, which can be described in two steps. First, a Sybyl atom type is assigned to each heavy atom of a molecule after removing the explicit hydrogen. Second, the individual atom fingerprint is calculated for every heavy atom by constructing a count vector that describes the neighbouring atoms from zero distance up to two bonds from the central atom. This second process creates features of atom environments, which are stored as a binary fingerprint of absence or presence of a feature. The second process is repeated for all assigned Sybyl atoms.

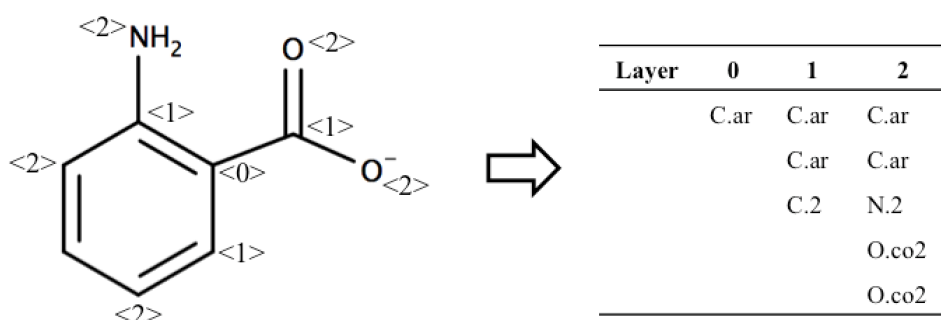


Figure 2.2: The figure illustrates an example of how a Molprint2D fingerprint is generated for an aromatic carbon, <0>, of a molecule. After assigning Sybyl atom type for each heavy atom of a molecule, the atom environment of the heavy atom is generated by describing its neighbouring atoms up two bonds (layers).

The targets predicted for the new compound uses the Naïve Bayes classifier as a method for classification as follows (105).

$$P(\omega_{\alpha}|\kappa) = \frac{P(\kappa|\omega_{\alpha})P(\omega_{\alpha})}{P(\kappa)}$$

(Equation 2.1)

Here, the probability of a new compound belonging to a target class, ω_{α} , with a given vector molecular feature, κ , is calculated. The prior target class probability, $P(\omega_{\alpha})$ is assumed to be equivalent to proportion of the molecules, $N\omega$, modulate that class from the total number of molecules, N , in the training set.

$$P(C = \omega) = \frac{N\omega_{\alpha}}{N}$$

(Equation 2.2)

The denominator, $P(\kappa)$ calculates the sum of the fraction of molecules from each class from $\beta = 1$ to L in the training set multiplied by the probability of the vector of molecular feature given the target class.

$$\sum_{\beta=1}^L \frac{N\omega_{\beta}}{N} P(\kappa|\omega_{\beta})$$

(Equation 2.3)

The posterior probability, $P(\kappa|\omega_{\alpha})$ is the likelihood of the feature, κ , given the class, ω_{α} .

The internal validation of the *in silico* target prediction was measured using 5-fold cross validation with a recall of correct targets that was larger than 80% in the top 1% of predictions. In the external validation, the algorithm showed a recall of 63.6% in the top 1% of predictions using dataset extracted from WOrld of Molecular BioAcTivity version 2011.1 (WOMBAT) (106). Only target scores above a confidence score, which are defined individually for each target class, were taken as the output. The confidence scores for each class was calculated by the optimal balanced accuracy (precision and recall trade-off) on a per target class basis, and were used to retain protein targets likely to interact with the compounds in the dataset (107).

2.2.3 Pathway annotation

Each predicted target was annotated with its full set of pathways from the KEGG biological

pathways (release 58.1) (108). It was possible to annotate 405 out of 477 targets with KEGG biological pathways (68, 109).

2.2.4 Enrichment calculations

To normalise the classification results from target prediction/pathway annotation, enrichment calculations were performed by normalising frequencies of the target prediction/pathway annotation of each therapeutic action subclass of compound to a background of 10,000 compounds that were selected randomly from PubChem (110) and ZINC (111), which consists of 194,849 compounds in total. Two methods were used to perform the enrichment calculation in this step, namely Estimation Score and Average Ratio. The calculation of the scores was performed as follows (68, 109).

1. Estimation Score

The Estimation Score is based on the frequency of the number of predicted targets/pathways in the random dataset, which is larger or equals the frequency of the number of predicted target/pathways in the test dataset. The absolute frequency, C , was divided by the total number of the random dataset, giving a value between 0 (enriched) and 1 (random).

$$Estimation\ Score = \frac{C}{1000}$$

(Equation 2.4)

2. Average Ratio

The Average Ratio is calculated by the ratio of the frequency (F) of the predicted target/annotated pathway in each random dataset with the frequency (F) of predicted targets in the test dataset.

$$Average\ Ratio = \frac{\frac{F(random\ set1)}{F(test\ set)} + \frac{F(random\ set2)}{F(test\ set)} + \dots + \frac{F(random\ set10000)}{F(test\ set)}}{10000}$$

(Equation 2.5)

In this study, enriched targets/pathways were considered if they showed an Estimation Score ≤ 0.01 , and descending Average Ratio was used to further discriminate important targets in agreement with previous work by Liggi et al. (68, 109). The relative cut-off of both Target Frequency (TF)/Pathway Frequency (PF) that was $\geq 5\%$ of the highest predicted target/pathway frequency was used after ranking the targets using the Estimation Score and Average Ratio methods to determine which targets were considered to be enriched in a particular therapeutic action (sub-)classes.

2.2.5 Hierarchical clustering based on the bioactivity space of the therapeutic action (sub-)classes

The frequencies of compounds across 477 targets for each therapeutic action (sub-)class were subjected to agglomerative hierarchical clustering (112, 113). The clustering method involved two steps as follows:

1) Selecting measures of dissimilarity or similarity

The dissimilarity distance between two therapeutic action (sub-)classes was calculated using the “dist” function of the ‘Euclidean’ method in R (114) after scaling the frequencies.

2) Clustering

Clustering was performed using the “hclust” function in R (114) based on the previously calculated Euclidean distance and Ward’s clustering method (115). In this method two clusters were merged if the sum of squared Euclidean distance was minimal.

A cut-off dissimilarity distance of approximately 20 was applied in order to obtain a manageable number of clusters, defining 14 groups of therapeutic action (sub-) classes, namely cluster I to XIV.

2.2.6 Targets and pathways analysis

The top three enriched targets/pathways were inspected with regard to their ability to explain the MOA of the compounds classified in the therapeutic action (sub-)classes. To improve the mechanistic understanding of the MOA, the top three enriched targets/pathways were linked to the indications of the (sub-)classes with supporting evidence from literature and supporting *in vitro* or *in vivo* studies of the Chinese medicines’ extracts or isolated compounds. However, supporting *in vitro* and *in vivo* studies were excluded in the pathway analysis because, in many (sub-)classes, no information was found. The 14 clusters were grouped based on the number of (sub-)classes in a cluster, which ranged from ten (sub-)classes to only one (sub-) class that derived from different classes. Three clusters were analysed in detail in the next section. Cluster VII was the only cluster that composed of (sub-)classes with the same TCM vital substance of its meridian system, blood. Cluster X was picked as a representative of a cluster of different classes while cluster XII was selected as a representative of a cluster with only one type of (sub-)class. The top three enriched targets/pathways per selected (sub-) classes were summarised in **Table 2.2** and **Table 2.5** and the top three enriched targets/pathways of all (sub-)classes can be found in **Table A2.1** and **Table A2.2**

(Appendices).

To compare the bioactivity spaces among the clusters, all enriched targets in a cluster were classified according to their protein family as derived from UniProt (116). The enriched pathways for all clusters were also classified according to KEGG ortholog groups, which are derived by comparing sequence similarity of individual genes and defining the functional group from the list of genes in the respective group (65). The KEGG ortholog group will be described as a pathway motif from here onwards. All the enriched targets were annotated from 59 protein families and the enriched pathways were annotated from 33 pathway motifs. A major protein family/pathway motif for a cluster was defined if the number of enriched targets/pathways was at least 5% of the total number of enriched target/pathway in the respective cluster and presence for at least eight clusters. Only five major protein families and eight major pathway motifs were identified. The frequencies of enriched targets/pathways per cluster were normalised before constructing two different heatmaps using heatmap.2 function of the gplots package in R (114), in order to visualise whether the major protein families/pathway motifs were equally important across clusters.

2.3 Results and Discussions

2.3.1 Overview of the results

The *in silico* target prediction of 10,749 TCM compounds yielded 409 unique targets, of which 183 were enriched targets. In the pathway annotation, the total number of unique pathways was 171, of which 99 were enriched pathways. The results discussed from here onwards cover 45 of the 46 therapeutic action (sub-)classes only. One therapeutic action subclass is not included because no target was retained from the “Tranquilizing - Settling” (Tranquil-settle) subclass, which contained only one compound. This subclass was therefore omitted from hierarchical clustering and all subsequent analysis. In **Figure 2.3**, a dendrogram shows the hierarchical clustering of 45 TCM therapeutic action (sub-)classes based on their bioactivity fingerprints. The cluster tree generates a diverse spread of the 45 therapeutic action (sub-)classes, which is defined into 14 clusters. In many instances, branches of the dendrogram paired up from different type of classes or subclasses. Based on the molecular similarity principle (41), this observation indicates that many similar compounds are present in both (sub-)classes despite having different therapeutic actions. The link of the top three enriched targets of the therapeutic action (sub-)classes are discussed in the next sections as well as their top three enriched pathways.

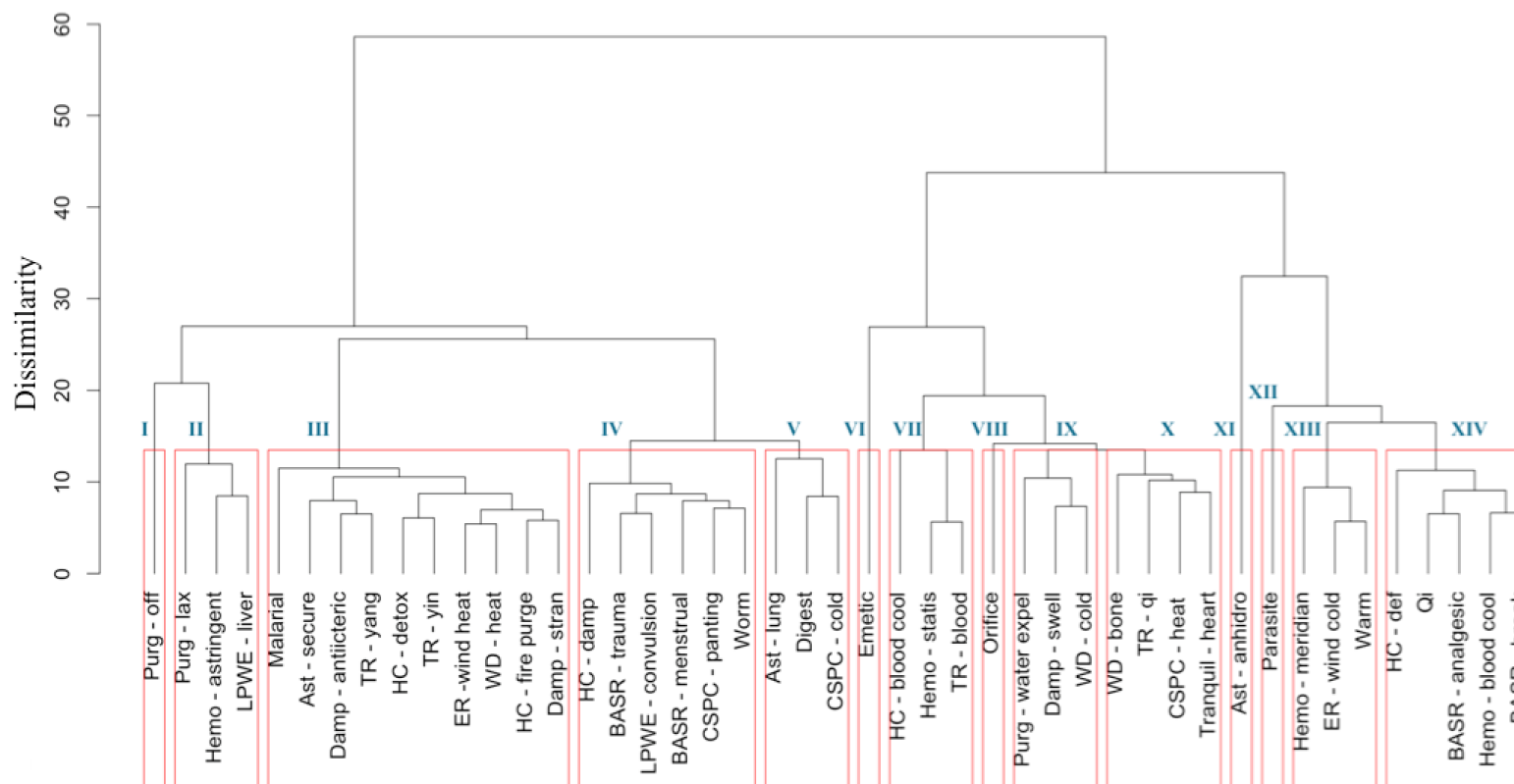


Figure 2.3: Hierarchical cluster analysis of TCM classes and subclasses is based on the similarity of the bioactivity fingerprint of each class. The ‘Tranquil-settle’ subclass was not included here (and in the further analysis) since it only contained a single compound for which no reliable targets could be predicted. The (sub-)classes were defined into 14 clusters, which cluster VII, X, and XII were selected for further analysis based on the top three enriched targets/pathways.

2.3.2 Target analysis

2.3.2.1 Clusters with four therapeutic action (sub-)classes (Cluster X)

Only Cluster X represents this group. The subclasses in cluster X are “Wind-dampness dispelling, bone (sinew) strengthening” (WD – bone), “Tonifying and replenishing, qi tonifying” (TR – qi), “Cough suppressing and panting-calming, clearing and heat phlegm resolving” (CSPC – heat) and “Tranquilizing, heat nourishing tranquilizing” (Tranquil – heart). The top three enriched targets from cluster X are mainly implicated in immunomodulation, namely steryl-sulfatase (STS) (117), tyrosine-protein phosphatase non-receptor type 2 (PTPN2) (118), and peptidyl-prolyl cis-trans isomerase FKBP1A (FKBP1A) (119), glucose homeostasis such as sodium glucose transporter 1 (SGLT1) and 2 (SGLT2) (120), cancer which is, DNA topoisomerase 1 (TOPO1) (121), reproductive system such as testosterone 17-beta-dehydrogenase 3 (17-beta-HSD 3) (122) and central nervous system (CNS) such as glutamate carboxypeptidase 2 (CGPII) (123).

Dissecting the *in silico* target prediction per subclass, starting from the “Wind-dampness dispelling, bone strengthening” (WD – bone), the top three enriched targets are TOPO1 (Estimation Score (ES)=0, Average Ratio (AR)=0.0144), SGLT1 (ES=0, AR=0.0342), and STS (ES=0, AR=0.0370). From the *in silico* target prediction, acankoreoside A-C derived from *Acanthopanax gracilistylus* were suggested to modulate TOPO1 and SGLT1 in the subclass, while compounds from *Homalomena occulta* such as asterpenoid, bullatantriol, and homalomenol were suggested to modulate STS, as well as compounds such as quercetin of *Taxillus chinensis*. The target prediction of compounds from *Acanthopanax gracilistylus* is supported by an *in vitro* study that the herb’s extract inhibits cell proliferation in several types of cancer cells (124). It is also reported that the extract of *Taxillus chinensis* exhibits significant anti-inflammatory activity *in vitro* (125). The herbs from this class are used to relieve pain, relax muscle and tendons, open channels and collaterals, and strengthen tendons and bones (87, 126). The actions of the herbs are related to the disturbance of muscular function in diabetes (127) and cancer cachexia which affects protein and lipid metabolism in skeletal muscle (128).

The top three enriched targets for “Tonifying and replenishing, qi tonifying” (TR – qi) subclass are PTPN2 (ES=0, AR=0.0174), SGLT2 (ES=0, AR=0.0282) and SGLT1 (ES=0, AR=0.0321). Many compounds from *Glycyrrhiza glabra*, *Glycyrrhiza uralensis*, *Dolichos lablab*, *Panax ginseng*, and *Astragalus membranaceus* were predicted to modulate PTPN2 and SGLT1. In particular to PTPN2 (129), a compound, glycyrrhizin from *Glycyrrhiza glabra* and *Glycyrrhiza uralensis* is supported by an *in-vivo* study that suggests the compound ameliorates all established chronic histopathologic changes of lung tissue in the mouse model

of asthma (130). TCM describes the medicines from this subclass act on the spleen and lung, which the deficiency of lung qi is characterised by shortness of breath such in asthma (87).

A sub-cluster consists of two therapeutic action subclasses is seen to have highly a similar bioactivity space. The first subclass is “Cough suppressing and panting-calming, clearing and heat phlegm resolving” (CSPC – heat) with the top three enriched targets being PTPN2 (ES=0, AR=0.0112), TOPO1 (ES=0, AR=0.0236), and 17-beta-HSD 3 (ES=0, AR=0.0341). Compounds from *Platycodon grandiflorum* such as platycoside A-M were suggested to be modulated by all the top three enriched targets and compounds from *Bambusa tuldoidea*, *Peucedanum decursivum*, and *Trichosanthes kirilowii* were mostly predicted to modulate 17-beta-HSD 3. Although *Platycodon grandiflorum* (131) and *Trichosanthes kirilowii* (132) are reported to exhibit anti-cancer properties, the reports do not support the link between the top three enriched targets and the indication of therapeutic action subclass to rationalise the MOA of the compounds.

The second subclass, “Tranquilizing, heart nourishing tranquilizing” (Tranquil – heart), lists FKB1A (ES=0, AR=0.0077), PTPN2 (ES=0, AR=0.0167) and CGPII (ES=0, AR=0.0209) as its top three enriched targets. Compounds of three different herbs were predicted to modulate FKB1A and PTPN2 and only compounds of *Ganoderma lucidum*, such as ganoderic acid and lucidenic acid derivatives were predicted to modulate CGPII. The prediction of CGPII from triterpenoids of *Ganoderma lucidum* such as ganoderic acids are supported by Zhang et al., which triterpenoids exhibit nerve growth factor or brain-derived neurotrophic factor activities *in vitro*, which has the therapeutic potential in neurodegenerative diseases (133). The therapeutic actions of the herbs from the subclass are described to have effects on central nervous system (87). In conclusion, the MOAs of the compounds for three subclasses in cluster X can be suggested from their top three enriched targets.

2.3.2.2 Clusters with three therapeutic action (sub-)classes (Cluster VII)

Cluster VII is one of the clusters that consists of three therapeutic action subclasses, with two subclasses, “Hemostatic, stasis resolving” (Hemo-stasis) and “Tonifying and replenishing, blood” (TR-blood) that shows a highly similar bioactivity space. The third subclass is “Heat-clearing, blood cooling” (HC-blood cool). Overall, the top three enriched targets in the cluster can be classified into immunomodulation which are PTPN2 (118), protein kinase C beta type (PKC- β) (134), protein kinase C eta type (PKC- η) (135) and protein kinase C gamma type (PKC- γ) (136), cancer, namely TOPO1 (121), and glucose homeostasis such as SGLT2 (120). The top three enriched targets in “Hemostatic, stasis resolving” (Hemo-stasis) and

“Tonifying and replenishing, blood” (TR-blood) are PTPN2 (ES=0, AR=0.0089), PKC- η (ES=0, AR=0.0182), PKC- γ (ES=0, AR=0.0209) and PTPN2 (ES=0, AR=0.0140), PKC- β (ES=0, AR=0.0230) and PKC- ϵ (ES=0, AR=0.0240), which all of them is implicated in immunomodulation. In the “Hemostatic, stasis resolving” (Hemo-stasis) subclass, compounds such as ginsenosides and notoginsenosides of *Panax notoginseng* were predicted to modulate all top three enriched targets. The target prediction also showed that anthraquinone compounds from *Rubia cordifolia* such as purpurin, ruberythric acid and soranjidiol modulate PKC- β and PKC- ϵ . In support of the target prediction of anthraquinone compounds of the second herb, a study reported the herb’s ethanol extract shows wound healing activities in mice, which from histological evaluations indicate a marked infiltration of the inflammatory cells, increasing blood vessel formation and enhancing cells proliferation (137). This finding agrees with the description of the subclass of which is to stop bleeding with a stabbing *pain* at a *fixed location* (87).

In the “Tonifying and replenishing, blood” (TR-blood) subclass, compounds from *Paeonia lactiflora* such as albiflorin, gallotannin, and casuarictin were predicted to modulate PKC- β and PKC- η . PTPN2 and PKC- η were both frequently predicted to be modulated by compounds from *Panax notoginseng* such as notoginsenoside and ginsenoside. It is found the ginsenoside Rg1 of *Panax notoginseng* ameliorates liver damage and suppresses proinflammatory cytokines secretion in concanavalin A-induced hepatitis in mice (138). This subclasses is described to have pharmacological effects on the liver, heart, and spleen and prevent failures of the organs (87). The top three enriched targets for “Heat-clearing, blood cooling” (HC-blood cool) subclass are PKC- β (ES=0, AR=0.0100), TOPO1 (ES=0, AR=0.0123) and SGLT2 (ES=0, AR=0.0137). In this subclass, compounds from *Paeonia lactiflora* such as albiflorin, isopaenoflorin, and benzoylpaenoflorin were predicted to modulate both PKC- β and TOPO1, while SGLT2 was predicted to be modulated by compounds from *Rehmannia glutinosa* such as rehmanionoside A and B. We found that the target prediction of SGLT2 is supported by a study on stachyose extract from *Rehmania glutinosa*, which shows a significant hypoglycaemic effect in diabetic mice (139). TCM views the action of the subclass is to promote the generation of body fluids from excessive heat (87, 126) which the consumption of body fluid is one of the symptom in diabetes (140) and cancer patients (141). Altogether, in many instances, the MOAs of the compounds can be explained from the enriched targets and can also be linked to the indications of the (sub-)classes.

2.3.2.3 Cluster with one therapeutic action (sub-)class (Cluster XII)

One out of five clusters, which has only one therapeutic action class is cluster XII, “Parasite destroying, dampness eliminating and itchiness relieving” (Parasite). The top three enriched targets are dihydrofolate reductase (DHFR), which plays a role in bacterial infection and cancer (142) (ES=0, AR= 0.0532), DNA-dependent protein kinase catalytic subunit (DNA-PKcs), which is implicated in cancer (143) (ES=0, AR=0.0644) and tumour necrosis factor (TNF), which is found to exert activities in cancer, inflammation, and bacterial infection (144, 145) (ES=0, AR=0.0687). From the target prediction compounds from *Allium sativum* were predicted to modulate DHFR, such as allicin, allilthiamine, and allyl disulphide. This prediction agrees with a study by Adetumbi et al., from which the extract of *Allium sativum* is found to inhibit the synthesis of proteins, nucleic acids, and lipids in *Candida albicans* where the major component of the herb was allicin (146). The finding relates to the phenotypes of the subclass, which is to kill and expel parasites and subsequently relieve pain (87, 126). In brief, the MOAs of the class can be linked to the enriched targets and the indications of the subclass. **Table 2.2** summarises the top three enriched targets of the therapeutic action (sub-)classes.

The above analysis can be summarised into two different views namely biological space and chemical space. From the biological space’s view, many of the top three enriched targets, regardless of therapeutic action (sub-)classes in any of the clusters, are implicated in immunomodulation such as PTPN2, PKCs, FKBP1A and STS, which PTPN2 and PKCs of which were frequently predicted. Both PKC family and PTPN2 are implicated in immunomodulation. The frequency of immune-related targets can be related when TCM influences the immune system regulation by either promoting or suppressing the immune factors (147). The PKC isoenzymes act as important mediators in immune cellular signalling in T- and B-lymphocytes in acquired immune system (148). PTPN2 also plays a major role in the transmission of immune cell signalling events (118). From the chemical space’s view, triterpenoid is the most frequent phytochemical (**Table 2.3**), which was predicted to modulate some of the top three enriched target, such as TOPO1, PTPN2, and PKCs. The compounds are found in different herbs such as *Acanthopanax gracilistylus*, *Glycyrrhiza glabra*, *Platycodon grandiflorum*, *Ganoderma lucidum* and *Panax notoginseng*. To validate the result from the chemogenomic principle that similar targets share similar compounds, a compound, ChEMBL 486881, from ChEMBL database (28) is most similar to acankoreoside A, with Tanimoto coefficient (TC) value of 0.56 (**Table 2.4**). Anti-proliferative activity against human HEK293 cells after 72 hrs by MTT conversion assay inhibition assay with reported $IC_{50} = 0.23\mu M$. Hence, the chemogenomics principle appears to stand (at least for acankoreoside A) because both compounds are implicated in cancer and structurally similar.

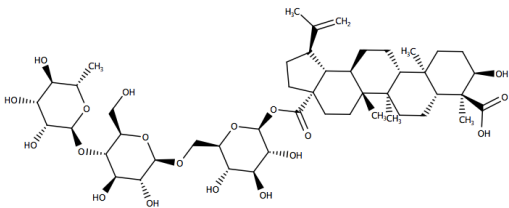
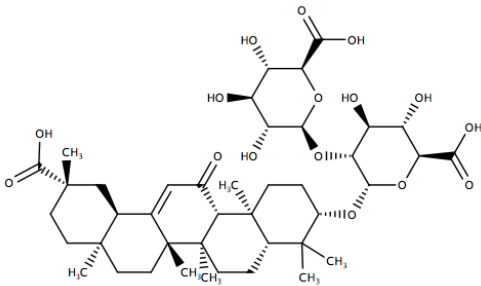
Table 2.2: The top three enriched targets in clusters VII, X, and XII. It can be seen that, in many cases, top three enriched targets are implicated in immunomodulation. The Estimation Score = 0, for all top three enriched targets.

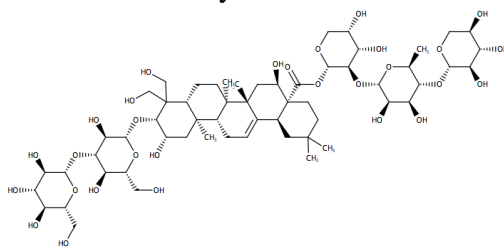
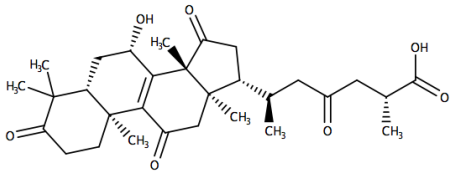
	TCM Therapeutic Action Class	TCM Therapeutic Action Subclass	Top three enriched targets	Target function reported by literatures	Average Score
Cluster X	Wind-dampness dispelling	Bone(sinew) strengthening	DNA topoisomerase 1	cancer	0.0144
			Sodium/glucose cotransporter 1	glucose homeostasis	0.0342
			Steryl-sulfatase	immunomodulation	0.0370
	Tonifying and replenishing	Qi tonifying	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation	0.0174
			Sodium/glucose cotransporter 2	glucose homeostasis	0.0282
			Sodium/glucose cotransporter 1	glucose homeostasis	0.0321
	Cough suppressing and panting-calming	Clearing and Heat phlegm resolving	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation	0.0112
			DNA topoisomerase 1	cancer	0.0236
			Testosterone 17-beta-dehydrogenase 3	reproduction system	0.0341
	Tranquilizing	Heat nourishing tranquilizing	Peptidyl-prolyl cis-trans isomerase FKBP1A	immunomodulation	0.0077
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation	0.0167
			Glutamate carboxypeptidase 2	CNS	0.0209
Cluster VII	Heat clearing medicinal	Heat clearing and blood cooling	Protein kinase C beta type	immunomodulation	0.0100
			DNA topoisomerase 1	cancer	0.0123
			Sodium/glucose cotransporter 2	glucose homeostasis	0.0137
	Hemostatic medicinal	Stasis-resolving hemostatic	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation	0.0089
			Protein kinase C eta type	immunomodulation	0.0182
			Protein kinase C gamma type	immunomodulation	0.0209
	Tonifying and replenishing	Blood tonifying	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation	0.0140

Cluster XII	Parasite destroying, dampness eliminating and itchiness relieving	Protein kinase C beta type	immunomodulation	0.0230
		Protein kinase C eta type	immunomodulation	0.0240
		Dihydrofolate reductase	cancer, bacterial infection	0.0532
		DNA-dependent protein kinase catalytic subunit	cancer	0.0644
		Tumour necrosis factor	cancer, bacterial infection	0.0687

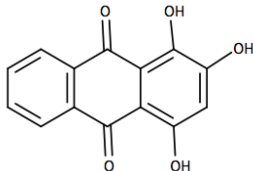
Table 2.3: Compounds from Clusters VII, X, and XII, which were found to modulate the top three enriched targets of the therapeutic action classes from the respective clusters. The phytochemical, triterpenoids were frequently found in different herbs. The predicted targets were supported by published literature based on at least one of the herbs, suggesting that the other herbs that also contain the same compound could contribute to related therapeutic effects.

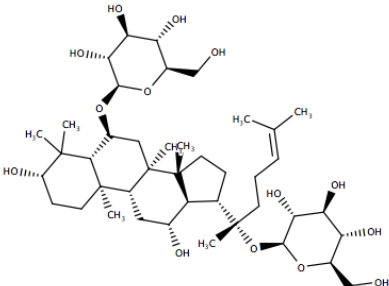
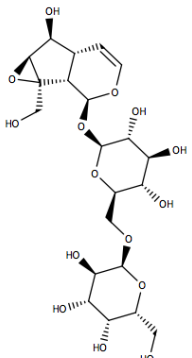
Cluster X

Therapeutic action (sub-class)	Herb	Compound	Predicted targets (based on the top three enriched targets)	Literature Support
“Wind-dampness dispelling, bone strengthening” (WD – bone)	<i>Acanthopanax gracilistylus</i>	Acankoreoside A 	TOPO1 SGLT1	The herb's extract inhibited cell proliferation several types of cancer cells (124).
Tonifying and replenishing, qi tonifying” (TR – qi)	<i>Glycyrrhiza glabra</i> <i>Glycyrrhiza uralensis</i>	Glycyrrhizin 	PTPN2	An <i>in vivo</i> study that suggested the compound ameliorates all established chronic histopathologic changes of lung in the mouse model of asthma (130).

Cough suppressing and panting-calming, clearing and heat phlegm resolving (CSPC – heat)	<i>Platycodon grandiflorum</i>	<p>Platycoside A</p> 	PTPN2 TOPO1 17-beta-HSD 3	No supporting literature.
Tranquilizing, heart nourishing tranquilizing (Tranquil – heart)	<i>Ganoderma lucidum</i>	<p>Ganoderic acid C1</p> 	CGPII	Triterpennoids of the herb exhibited nerve growth factor or brain-derived neurotrophic factor activities <i>in vitro</i> , which has the therapeutic potential in neurodegenerative diseases (133).

Cluster VII

Therapeutic action (sub-class)	Herb	Compound	Predicted targets (based on the top three enriched targets)	Literature Support
Hemostatic, stasis resolving (Hemo-stasis)	<i>Rubia cordifolia</i>	<p>Purpurin</p> 	PKC- β PKC- ϵ	The ethanol extract showed wound healing activities in mice, which from histological evaluations indicated a marked infiltration of the inflammatory cells, increased blood vessel formation and enhanced proliferation of cells (137).

Tonifying and replenishing, blood (TR-blood)	<i>Panax notoginseng</i>	<p>Ginsenoside Rg1</p> 	PTPN2 PKC- η	The compound ameliorated liver damage and suppressed proinflammatory cytokines secretion in concanavalin A-induced hepatitis in mice (138).
“Heat-clearing, blood cooling” (HC-blood cool)	<i>Rehmania glutinosa</i>	<p>Rehmannioside A</p> 	SGLT2	The stachyose extract from the herb showed a significant hypoglycaemic effect in diabetic mice (139).

Cluster XII

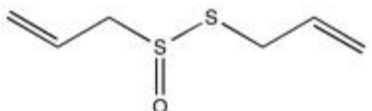
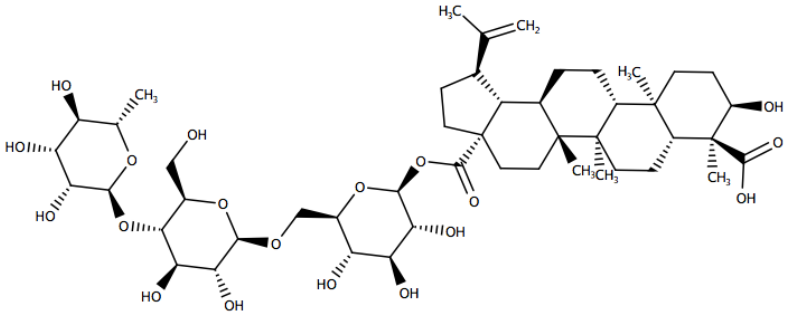
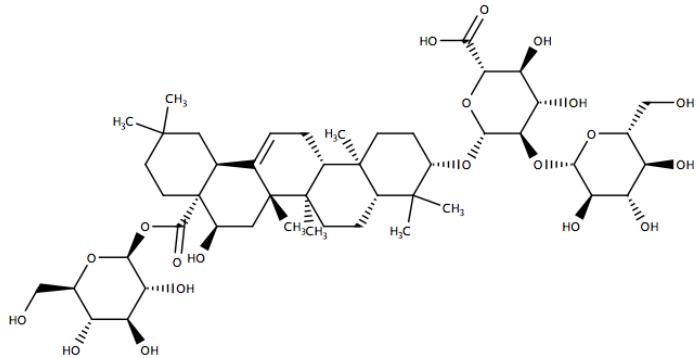
Therapeutic action (sub-class)	Herb	Compound	Predicted targets (based on the top three enriched targets)	Literature Support
Parasite destroying, dampness eliminating and itchiness relieving (Parasite)	<i>Allium sativum</i>	<p>Allicin</p> 	DHFR	The extract of <i>Allium sativum</i> was found to inhibit the synthesis of proteins, nucleic acids, and lipids in <i>Candida albicans</i> where the major component of the herb was allicin (146).

Table 2.4: Compound from ChEMBL database that is most similar to acankoreoside A (**Table 2.3**) and Its Activity Profile

Reference compound	Closest Similarity	Reported Activity Profile
<p style="text-align: center;">Acankoreoside A</p> 	<p style="text-align: center;">CHEMBL 486881</p> 	<p>Anti-proliferative activity against human HEK293 cells after 72 hrs by MTT conversion assay inhibition assay with reported $IC_{50} = 0.23\mu M$</p> <p><i>Tanimoto coefficient</i> = 0.56</p>

2.3.3 Pathway analysis

2.3.3.1 Clusters with four therapeutic action (sub-)classes (Cluster X)

The top three enriched pathways in cluster X are mainly implicated in digestive system (carbohydrate digestion and absorption, bile secretion and mineral absorption). In this cluster, only one pathway, from the top three enriched pathways, is not classified in the digestive system, which is terpenoid backbone biosynthesis.

In the “Wind-dampness dispelling, bone strengthening” (WD – bone) subclass, the top three enriched pathways are mineral absorption (ES=0, AR=0.0342), carbohydrate digestion and absorption (ES=0, AR=0.1427) and bile secretion (ES=0, AR=0.2050). To link the mineral absorption pathway to the indication of the subclass, it has been reported that minerals such as calcium is an important mineral for bone remodelling to strengthen the bone (149). In the second pathway, carbohydrate digestion and absorption, is associated to the indication of the subclass, when a study shows that a lower rate of mesenchymal stem cells differentiated into osteoblasts when mice were fed with low carbohydrate - high fat (LC-HF) diets, which explains the observation of reduced bone formation (150). However, no study can be found to link bile secretion to the indication of the subclass.

“Tonifying and replenishing, qi tonifying” (TR – qi) subclass has mineral absorption (ES=0, AR=0.0321), carbohydrate digestion and absorption (ES=0, AR=0.1251) and bile secretion (ES=0, AR=0.2678) in the top three enriched pathways. Minerals play vital roles in maintaining the cell functions to optimise health and prevent diseases. The first enriched pathway, mineral absorption, the process of recycling iron in erythrocytes is seen in splenic macrophages in the red pulp (151) and agrees with the indication of the subclass which is qi tonifying related to the maintenance of blood flow within the vessels that is implicated directly from spleen’s activity (126). In the second pathway, no study can be found to link carbohydrate digestion and absorption to the indication of the subclass. Relating to the third enriched pathway, bile secretion, it has been reported that the major components of bile acid, chemodeoxycholic acid and glycochenodeoxycholic acid can induce cyclooxygenase-2 expression and cell proliferation in esophageal squamous cells, suggesting that bile acids may contribute to the inflammation and mucosal thickening (152). Another paper demonstrates that bile acids may induce airway fibrosis through the production of TGF- β 1 and fibroblast proliferation (153). These findings can be associated with the indication of the subclass, which is related to lung deficiency (87).

The top enriched pathways of “Cough suppressing and panting-calming, clearing and heat phlegm resolving” (CSPC – heat) subclass are minerals absorption (ES=0, AR=0.0380), carbohydrate digestion and absorption (ES=0, AR=0.1572) and bile secretion (ES=0,

AR=0.1958). Relating to the first enriched pathway, mineral absorption, the abnormal distribution of trace elements such as zinc, selenium and copper have been reported to aggravate oxidative damage and inflammation in the airways and subsequently decreased the lung's function in asthmatic patient (154), in which an asthmatic condition is described as retention of heat phlegm in the lung (155). In the second enriched pathway, carbohydrate digestion and absorption, the ingestion of carbohydrate is reported to attenuate the migration of T lymphocytes to the bronchial epithelial cell line when it is infected with the common respiratory pathogen human rhinovirus during strenuous exercise (156). This effect agrees with the indication of the subclass which is to dissolve phlegm upon infection in the lung (87, 126). However, no strong evidence can be found to link bile secretion to the indication of the subclass.

The “Tranquilizing, heart nourishing tranquilizing” (Tranquil – heart) subclass lists mineral absorption (ES=0, AR=0.0387), carbohydrate digestion and absorption (ES=0, AR=0.1163) and terpenoid backbone biosynthesis (ES=0, AR=0.2166). In the first enriched pathway, mineral absorption, it is reported that selenium plays an important role in the brain where its deficiency is implicated in senility and Alzheimer's disease (157). In the third enriched pathway, terpenoid backbone biosynthesis includes mevalonate and non-mevalonate pathway. The mevalonate pathway is a pathway implicated in cholesterol biosynthesis in the brain and deficiencies in cholesterol metabolism can lead to diseases of the central nervous system (CNS) diseases (158). Both findings agree with the indication of the subclass, which is the pharmacological effects are on the central nervous system (87). Although no strong evidence to support, the second enriched pathway, carbohydrate digestion and absorption, the two enriched pathways can be linked to the indication of the subclass. In short, in many instances, the top three enriched pathways can be associated to the indications of the (sub-)classes to explain their MOAs.

2.3.3.2 Cluster with three therapeutic action (sub-)classes (Cluster VII)

The top three enriched pathways in the cluster can be classified into digestive system (carbohydrate digestion and absorption, bile secretion and mineral absorption), cellular communication (tight junction) and membrane transport (ABC transporters). The top three enriched targets in “Hemostatic, stasis resolving” (Hemo-stasis) subclass are mineral absorption (ES=0, AR=0.0267), carbohydrate digestion and absorption (ES=0, AR=0.0987) and bile secretion (ES=0, AR=0.1872) and “Tonifying and replenishing, blood” (TR-blood) subclass lists mineral absorption (ES=0, AR=0.0329), carbohydrate digestion and absorption (ES=0, AR=0.1055) and tight junction (ES=0, AR=0.2053). In “Hemostatic, stasis resolving”

(Hemo-stasis) subclass, the mineral absorption pathway can be related to zinc deficiency, which has been reported to delay wound healing (159). In the bile secretion pathway, bile acids are implicated in platelet inhibition by solubilising the platelets, which patients with obstructive jaundice were exposed to abnormal hemostasis due to high level of bile acids (160). Both enriched pathways are related to the indication of the subclass, which is to achieve hemostasis (87) and no study can be found to associate carbohydrate digestion and absorption pathway to the indication of the subclass. In “Tonifying and replenishing, blood” (TR-blood), the first enriched pathway, mineral absorption, can be linked to the indication of the subclass which is to strengthen the heart that controls blood vessel (87), when selenium is reported to be protective against cardiovascular disease by contributing to the production of vasodilatory prostacyclin by the endothelium (157). No study can be found to support the link between the second pathway, carbohydrate digestion and absorption to the indication of the subclass. The third enriched pathway, tight junction, the presence of tight junctions in the bile epitheliums act as barriers from toxic diffusion from bile into hepatic interstitial tissue, which could impair the organ’s function(161). TCM enriched can be associated with the indication of the subclass, which is to strengthen the function of liver (87).

The top three enriched targets for “Heat-clearing, blood cooling” (HC-blood cool) subclass are mineral absorption (ES=0, AR=0.0276), carbohydrate digestion and absorption (ES=0, AR=0.0753) and ABC transporter (ES=0, AR=0.0769). The mineral absorption pathway can be linked to the subclass when a decreased in intracellular magnesium concentration is implicated in type 2 diabetes (162) and the disease, according to TCM, is described as the deficiency of body liquid due to heat syndrome (163). Similar to previous subclass, no study can be found to link the second enriched pathway, the carbohydrate digestion and absorption, to the indication of the subclass. In the third pathway, ABC transporter, it is reported that the high expression of ABCG5 and ABG8 in hypercholesteromic condition of the heart is involved in cardiovascular protection by lowering plasma cholesterol level (164). The pathway can be linked to the indication of the subclass when the herbs are described to act on liver and heart (87). All in all, in many cases, the top three enriched pathways can be linked to the indications of the subclasses to explain their MOAs.

2.3.3.3 Cluster with one therapeutic action (sub-)classes (Cluster XII)

In the “Parasite destroying, dampness eliminating and itchiness relieving” (Parasite) class, the top three enriched pathways are steroid biosynthesis (ES=0, AR=0.2530), glycerophospholipid metabolism (ES=0, AR=0.2941) and collecting duct acid secretion

(ES=0, AR=0.3420). Both of the first two pathways, steroid biosynthesis and glycerophospholipid metabolism are part of lipid metabolism, and many reports have suggested their link to immune response. For instance, steroid biosynthesis is down-regulated by interferon type I upon viral infection (165) and the chlamydia exploits the nutrient-rich host cell cytosol by trafficking the glycerophospholipid from the host cell for survival (166). In relation to the collecting duct acid secretion pathway, bacterial infection in the kidney is reported to affect the collecting duct acid secretion because the presence of lipopolysaccharide (LPS) of the bacteria inhibits the HCO_3^- absorption (167). These pathways agree with the functions of the subclass, which is to kill and expel parasites (87, 126). Altogether, the top enriched pathways can be linked to the indication of the class to explain the MOAs. **Table 2.5** summarises the top three enriched pathways of the therapeutic action (sub-)classes.

The above analysis can be summarised into three observations as follows. First, a type therapeutic action (sub-)class is implicated in different enriched pathways and each pathway is involved in a different disease. Second, more than one therapeutic action (sub-)classes are implicated in a pathway that is involved in multiple diseases. Third, more than one therapeutic action (sub-)classes are implicated in different pathways but involved in only one type of a disease. The first observation is indicated, in many cases, by the (sub-)classes in any of the three clusters such as bile secretion and mineral absorption in “Tonifying and replenishing” (TR-qi). The top three enriched pathways in a therapeutic action (sub-)class are implicated in different pathways and diseases. The modulation of one pathway to a disease could provide a better insight on the (sub-)class MOA in the biological system. In the second observation, the mineral absorption in cluster VII and cluster X and carbohydrate digestion and absorption in cluster VII are implicated in various types of diseases. The redundancy of a pathway in the pathogenesis of various diseases implies that a pathway could serve multiple purposes; for instance, the mitogen-activated protein kinase (MAPK) signalling pathway was implicated in inflammation, cancer, cardiovascular dysfunction and Alzheimer’s disease (168). The third observation shows that different pathways from different therapeutic action subclasses in a cluster, are implicated in a type of physiological function, such as the ABC transporter and mineral absorption from “Heat-clearing, blood cooling” (HC-blood cool) and “Tonifying and replenishing, blood” (TR-blood) respectively is implicated in physiological function of the heart. The different pathways for a particular physiological function explain that a pathway does not function alone in the manifestation of a disease but through the interactions of multiple pathways (169), which account for the different clinical symptoms. To put it briefly, the pathways annotation is not only beneficial to suggest the MOAs of the (sub-)classes based

on their indications, but also show that one pathway could have manifold functions and multiple pathways contribute to the pathogenesis of a disease. From the TCM perspective, it is suggested that the involvement of multiple pathways in the pathogenesis of a disease explains the complex TCM formulation, which consist of a set of herbs from various therapeutic action (sub-)classes.

Table 2.5: The top three enriched pathways in clusters VII, X, XII. It can be seen that, in many cases, similar pathways appear in the top three enriched targets regardless of clusters and subclasses. The Estimation Score = 0, for all top three enriched pathways.

	TCM Therapeutic Action Class	TCM Therapeutic Action Subclass	Top three enriched pathways	Average Score
Cluster X	Wind-dampness dispelling	Bone (sinew) strengthening	hsa04978 Mineral absorption	0.0342
			hsa04973 Carbohydrate digestion and absorption	0.1427
			hsa04976 Bile secretion	0.2050
	Tonifying and replenishing	Qi tonifying	hsa04978 Mineral absorption	0.0321
			hsa04973 Carbohydrate digestion and absorption	0.1251
			hsa04976 Bile secretion	0.2678
	Cough suppressing and panting-calming	Clearing and Heat phlegm resolving	hsa04978 Mineral absorption	0.0380
			hsa04973 Carbohydrate digestion and absorption	0.1572
			hsa04976 Bile secretion	0.1958
	Tranquilizing	Heat nourishing tranquilizing	hsa04978 Mineral absorption	0.0387
			hsa04973 Carbohydrate digestion and absorption	0.1163
			hsa00900 Terpenoid backbone biosynthesis	0.2166
Cluster VII	Heat clearing medicinal	Heat clearing and blood cooling	hsa04978 Mineral absorption	0.0276
			hsa04973 Carbohydrate digestion and absorption	0.0753
			hsa02010 ABC transporters	0.0769
	Hemostatic medicinal	Stasis-resolving hemostatic	hsa04978 Mineral absorption	0.0267
			hsa04973 Carbohydrate digestion and absorption	0.0987
			hsa04976 Bile secretion	0.1872
	Tonifying and replenishing	Blood tonifying	hsa04978 Mineral absorption	0.0329
			hsa04973 Carbohydrate digestion and absorption	0.1055
			hsa04530 Tight junction	0.2053
Cluster XII	Parasite destroying, dampness eliminating and itchiness relieving		hsa00100 Steroid biosynthesis	0.253
			hsa00564 Glycerophospholipid metabolism	0.294
			hsa04966 Collecting duct acid secretion	0.342

2.3.4 Comparison of bioactivity spaces of clusters

In this part of study, the aim was to investigate the differences of bioactivity spaces among all clusters by classifying all the enriched targets in the cluster to their respective protein families. The 181 enriched targets were classified into 59 protein families. Out of 59 protein families, five protein families were frequently annotated in all clusters, which are GPCR, protein kinase, NHR, carbonic anhydrase and cytochrome P450. The heatmap in **Figure 2.4** compares the five major protein families that were annotated based on the enriched targets in each cluster. The more saturated colour represents the more significant the protein family across all clusters. The numbers of protein families in the clusters were normalised because the distribution of enriched targets were not consistent due to numbers of therapeutic action (sub-)classes per cluster were different.

What can be seen from the graph is that GPCRs and protein kinases are the most highly clustered protein families in almost all clusters. The results were expected as both of the families are the two most largest protein families that are involved in many physiological processes (170, 171), thus, explaining why these two protein families were observed to be significant in many of the clusters. In addition, the diversity of the compounds in the (sub-)classes also contributed to the prediction of enriched targets from these protein families. At least four protein families were significant across all clusters, except for cluster VII and X in cytochrome P450. The number of enriched targets in the cluster was amongst the lowest across all clusters, cytochrome P450 were less classified. NHRs were found to be annotated in all clusters and the frequent predictions of the nuclear hormone receptor family can be explained by the presence of naturally occurring steroids in natural compounds (172). For the remaining two major protein families, the cytochrome P450 family was also expected to be modulated by most of the subclasses because this protein family plays an important role in the degradation of structurally rather diverse exogenous compounds (173). The carbonic anhydrase family, which is an ubiquitous enzyme, involved in the inter-conversion between carbon dioxide and the bicarbonate ion that is important for many physiological processes (174). All in all, five major protein families were observed to occur frequently in most of the clusters and were heavily implicated in the biological processes such as cell regulation, sensory system and steroid metabolism. This analysis has allowed the discovery of the bioactivity space connection between subclasses in TCM based on the sets of enriched targets from our *in silico* target prediction. The TCM compounds of a significant cluster for a particular protein family can be further explored in finding new molecular entities for a disease related to the protein family.

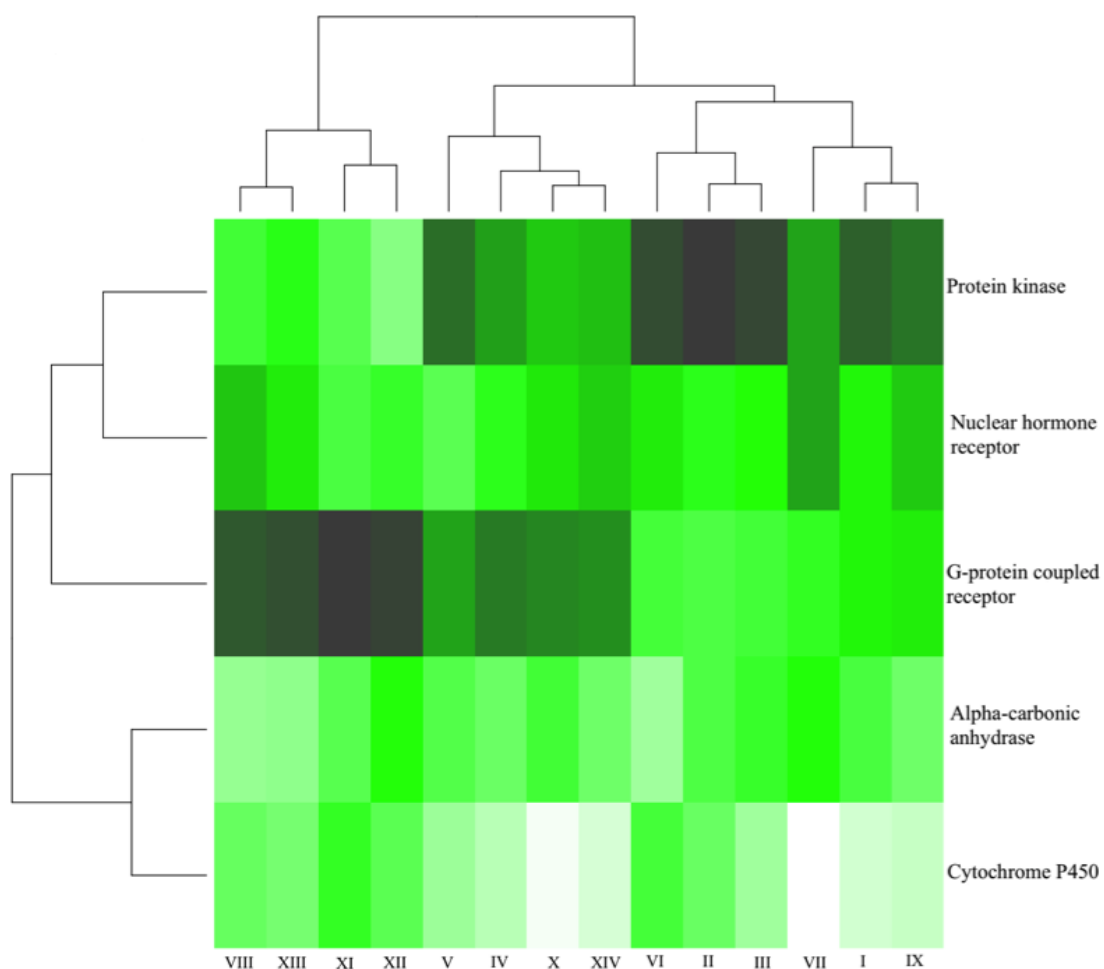


Figure 2.4: The heatmap compares the five major protein families that were annotated based on the enriched targets in each cluster, which were normalized. The more saturated colour across clusters represents the more significant the protein family. GPCR and protein kinase are observed to be significant protein families in almost all clusters. It appears that all the protein families are heavily implicated in the biological processes such as cell regulation, sensory system and steroid metabolism. The TCM compounds of significant cluster for a particular protein family can be suggested to be further explored for finding new molecular entities.

2.3.5 Comparison of pathway annotation of clusters

In this part of the study, we aimed to investigate the differences in pathway motifs among all clusters by classifying the enriched pathways according to KEGG ortholog. The 99 enriched pathways were classified to 33 pathway motifs, which were almost half the number of total pathway motifs available in KEGG. The major pathway motifs from the classification were infectious diseases, digestive system, immune system, signal transduction, lipid metabolism, cancer and cellular community. The heatmap in **Figure 2.5** compares the seven major pathway motifs that were annotated based on the enriched pathways in each cluster. The more saturated colour across clusters represents the more significant the pathway motif. The numbers of pathway motifs in the clusters were normalised due to the differences in numbers of the pathways that were enriched among all (sub-)classes

Figure 2.5, it is seen that the digestive system is consistently classified in all clusters, which in many cases, the plots' colour are more saturated compare to other pathway motifs. The digestive system includes the digestion and absorption of macro- and micronutrients and shows that the majority of the enriched pathways are bile secretion, pancreatic secretion and gastric acid secretion. The bile secretion controls the cholesterol homeostasis by routing the elimination of cholesterol, in addition to harmful exogenous lipophilic substance (175). In addition to the digestive system, infectious diseases, signal transduction, and lipid metabolism are significant pathway motifs in many of the clusters. The significance of infectious diseases can be deduced from frequently predicted immunomodulatory targets as the pathogenic factors are described to attack weakened immune system (176). The signal transduction pathway motif can be deduced from the classification of one of the major protein families, GPCR which translates the extracellular signals for the downstream effectors that produce a physiological response in a target cell (177). Many Chinese medicines have been reported to have lipid regulating effects by influencing the intestinal lipid absorption and lipid metabolism, to name a few (178). Also, one of the highly observed enriched pathways in the lipid metabolism pathway motif was steroid hormone biosynthesis, in which the cytochrome P450 protein family is involved in (179). It also appears that a few of the pathway motifs are insignificant for some clusters such as the immune system for cluster II and the excretory system in cluster XIV, which resulted from a low number of compounds influencing the target prediction and pathway annotation. The list of compounds that were annotated for the Chinese medicines in the subclasses might be incomplete in order to influence the classification of the immune system as well as the remaining missing pathway motifs. In addition, only enriched pathways were used in the classification. In a similar manner, for the major protein families, the compounds from

significant clusters for the major pathway motifs can be further explored for types of diseases that are well known to be implicated such as the digestive system, which is well known to be involved in liver disease.

Altogether, this analysis has allowed the discovery of major pathway motifs in all clusters. Despite having the different therapeutic action (sub-)classes in a cluster, in many cases, all the clusters can be classified of having the seven major pathway motifs. The classification of the major pathway motifs was associated with the major protein families analysed in the previous section.

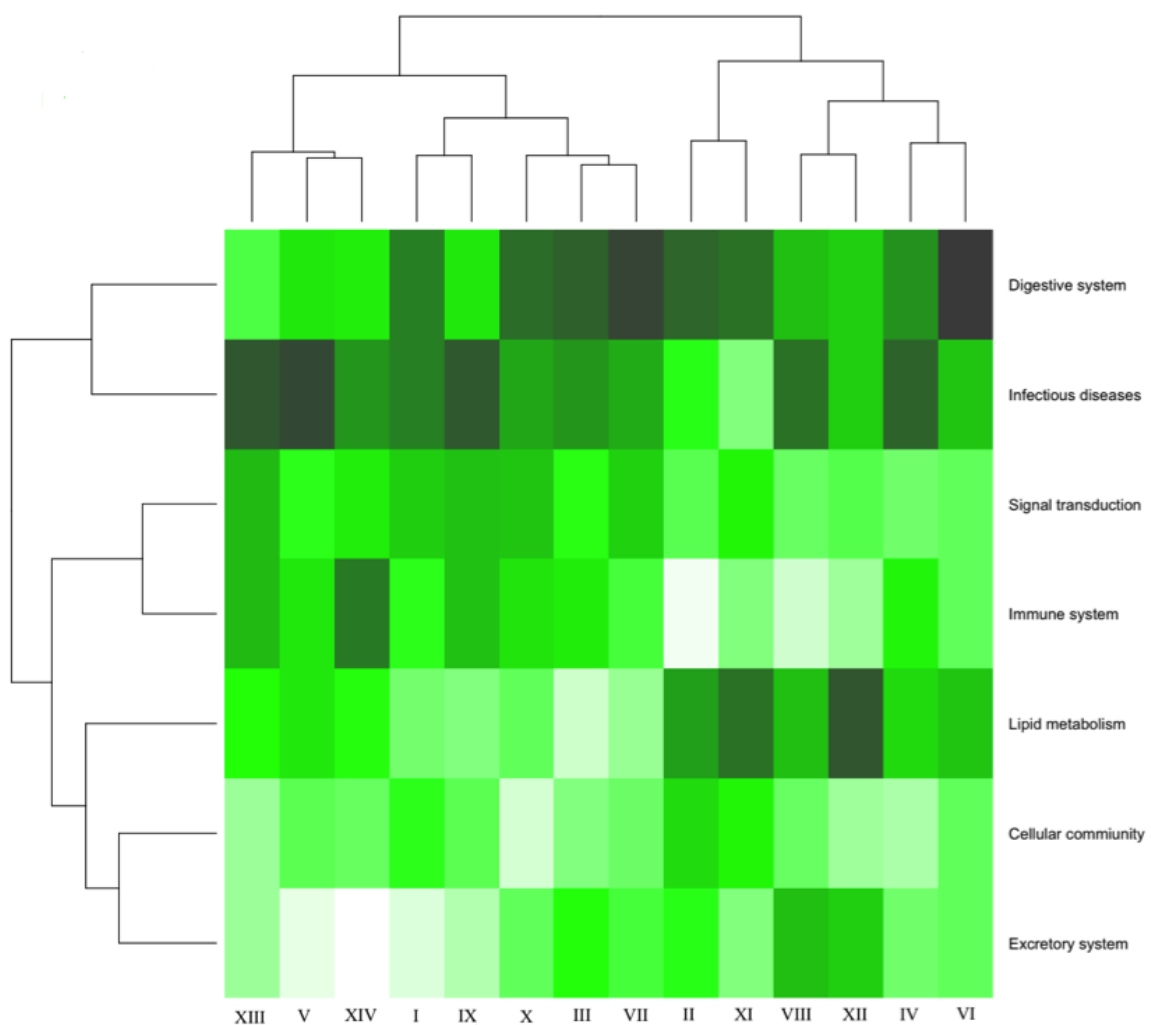


Figure 2.5: The heatmap compares the seven major pathway motifs that were annotated based on the number of enriched pathways in each cluster, which were normalized. The more saturated colour across clusters represents the more significant pathway motifs. The significant cluster for a particular pathway can be further explored for a disease with the known pathway motif such as digestive system in liver disease.

2.4 Conclusion

The global mapping of relationships between TCM therapeutic action classes and subclasses, based on their predicted targets and annotated pathways, provides a novel approach to understand the MOAs of TCM formulations. The dendrogram, which was generated based on the *in silico* target prediction of TCM therapeutic action (sub-)classes compounds, has enabled the visualisation of their bioactivity space. In the first part of the study, it was to rationalise the link between the top three enriched targets/pathways to the description of the respective therapeutic action (sub-)classes with supporting literature. Overall, the most frequent top three enriched targets were immune-related targets such as tyrosine-protein PTPN2 and PKC-family, while the most frequent enriched pathway was related to the digestive system such as mineral absorption and bile secretion. In TCM, symptoms are usually regarded as the invasion of pathogenic factors, thus sensitizing the immune system to response, and this might provide a mechanistic link between TCM and Western thinking. In the second part of the study, the annotation of the protein family in all clusters showed that the GPCR and protein kinase family were the two major protein families that contributed to the diversity of the bioactivity space. The highly annotated pathway motif indicated that the digestive system was consistently annotated, which agreed with the important treatment principle of TCM, “the foundation of acquired constitution” that includes spleen and stomach. With the global overview of the bioactivity space of the therapeutic action (sub-)classes, the similarity could be observed and the differences between them, which are not apparent from the name given to the (sub-)class itself. Hence, this analysis helps to reduce the gap between TCM and Western medicine.

Chapter 3:

Exploring the Chemical Space and Bioactivity Space of Traditional Medicine from China, India and Malaysia for Treating Cancers

3.1 Introduction

Traditional medicines (TMs) from different geographical locations are often based on different ingredients and formulations; however, the diseases to be treated are in many cases related. In many TM practices, plants are usually the major resource but some animal parts and minerals are also frequently used. Each TM can use different types of plants (which is an aspect we will investigate further in the current work) to treat the same diseases such as infectious diseases (*180*), memory and cognitive function (*181*) and cancer (*182*). These different plants do not only have structurally different compounds but also similar compounds. These compounds, most of which exhibit pharmacological properties, are known as secondary metabolites. The distribution of similar compounds in different plants can be explained by two notions. First, the plants are closely related in the phylogenetic tree, thus, contributing to the distribution of similar compounds (*183*). The relationship between phylogeny and shared phytochemistry has been reported before (*184, 185*); it is described as the characteristics of the plants' taxa despite being found in different regions. Second, the commonality, paradoxically, can also be explained by convergent evolution, where the production of secondary metabolites evolves independently in different plant phylogenetic lineages, in response to similar environmental challenges (*186, 187*). The occurrence of similar compounds from plants of unrelated taxa was found in a few studies (*97, 188*), which indicates the importance of a few types of secondary metabolites in plants defence mechanisms being exploited for medicinal purposes. Thus, it can be seen that structurally similar compounds can be found in either the same or unrelated niches of the phylogenetic tree of plants used to treat the same diseases.

The variability of plants used in treating the same disease can also be attributed to different compounds modulating different related targets or different genetic background to respond to TM treatment. Knowledge of the therapeutic target may help uncover the MOA of TM compounds and how the genes contribute to the development of the disease. Hence, the common and differential MOAs of different TMs could be understood better by comparing their chemical and bioactivity spaces. Thus far, investigations to compare the bioactivity spaces of different TMs are still limited (*180, 182*), where both studies included different TMs only from within the same region. Unlike the bioactivity space, extensive studies have investigated the variability of the chemical space between TMs but in comparison to

combinatorial libraries (189, 190). Hence, previous comparison studies were either restricted to only chemical or target space, whereas we aim to compare them concurrently.

To understand the relations between the likely MOAs of TMs, we used *in silico* target prediction. While previous studies have investigated links between TCM and Ayurveda and their putative targets on a large scale (47), in this work we explored the links between three TMs and their putative targets in cancer, which is one of the diseases treated by many TMs and where sufficient data for analysis is available.

Cancer was probably first defined by the physician Hippocrates (460–370 B.C) from a Greek word, *karkinos*, to describe carcinoma tumours (191). However, communities from different parts of the world had established their own philosophy of treatments to fight against cancer using TM long before the disease was defined. The three types of TM, namely Traditional Chinese Medicine (TCM), Ayurveda, and Malay TM, hold distinct views about cancer (**Figure 3.1**). According to TCM, cancer occurs when there is an imbalance between endogenous physical conditions, which is associated with the dysfunction of viscera and bowels, particularly the spleen and kidney deficiencies of the body and also the interference of exogenous pathogenic factors (141, 192). Cancer can be briefly concluded as the stagnation of toxin and heat, the obstruction of phlegm/dampness, Qi stagnation and blood stasis, and the imbalance of yin and yang in the viscera and bowels (192). On the other hand, Ayurveda believes that cancer is due to the disruption of aura that allows negative astral forces to enter the body (193). Cancer is described as inflammatory or non-inflammatory swelling which can either be a minor (*Granthi*) or major (*Arbuda*) neoplasm (194, 195). The swelling starts off when one or two of the three bodily systems, which are the nervous system (*Vata* or air), the venous system (*Pitta* or fire), and the arterial system (*Kapha* or water), which maintain the normal function of the body, are out of control in benign cancer and all of them become dysfunctional in malignant cancers due to clashes between *Vata* and *Kapha* forces resulting in morbidity (193, 194). In Malay TM, the physical characteristics of a person consist of four elements, which are fire (*Suprawi*), earth (*Suddawi*), wind (*Dammawi*) and water (*Balpawi*), by which a disease can be identified by recognizing the nature of the elements; damp, cold, hot and dry (196). Limited knowledge has been documented on how the Malay TM views cancer, but the symptom of cancer is described as a swelling that is due to the imbalance in the wind, which affects how the blood flows (197). Therefore, altogether, the above TMs agree that cancer is viewed as an obstruction in the normal body function, within the particular reference frame of each TM, which is described in modern medicine as the uncontrolled division of abnormal cells, that also will eventually disrupt the body function (191).

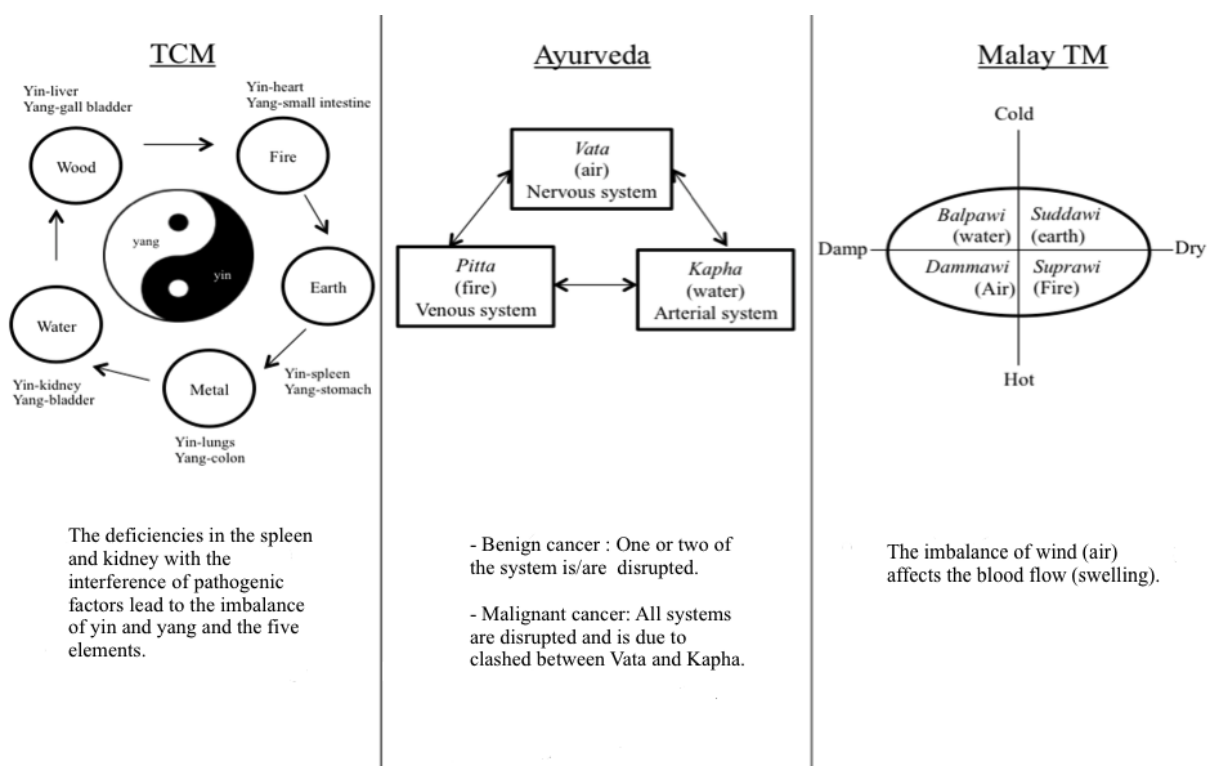


Figure 3.1: The comparison of the cancer definitions from TCM, Ayurveda, and Malay TM. In principle, each TM defines a healthy body through different elements. One of the elements is viewed to be disrupted when cancer develops. In TCM, the imbalance of yin and yang from spleen and kidney deficiencies are described as symptoms of cancer. Ayurveda believes that one of the three elements that holds a healthy body is disrupted in cancer. Similar in Malay TM, cancer happens when there is an imbalance of the wind element, which is a part of the four elements used to describe as state of a healthy body.

This current study aims to undertake the analysis of predicted targets differently by comparatively exploring both chemical and target space of the three TMs, i.e. TCM, Ayurveda and Malay TM with respect to treating cancer. The chemical space of TM compounds was compared with anti-cancer drugs in addition to diversity analysis of different TMs by scaffold decomposition. TM compounds were predicted using an *in silico* target prediction algorithm and the bioactivity space between TMs was compared by analysing the relation between predicted targets and their annotated plants and the significant target class of a TM. A phylogenetic tree was also constructed to determine its influence in the comparison study. In the end, this study allowed us to better understand the relations between TMs in treating cancer by analysing their common and differential MOAs.

3.2 Materials and methods

3.2.1 Compilation of traditional medicine datasets

Three TM datasets were used for this study. The TCM (97) and Ayurveda (193, 194, 198) plants used for cancer treatments (cancer-related plants) were extracted from literature. All TCM compound structures were collected from the Traditional Chinese Medicine Systems Pharmacology Database Platform (TCMSP) database (32) by searching the plants' scientific names. Out of 97 plants in (97), only 47 plants were found in the TCMSP. Due to limited Ayurveda databases available, the set of compounds of each Ayurveda plant was derived from the Dr. Duke's Phytochemical and Ethnobotanical Database (199) and the corresponding compound structures were downloaded from PubChem (110) using PubChem IDs as input in the PubChem Download Service. Where the structure was not available in PubChem (110), ChemSpider (25) or the Human Metabolome Data (HMDB) (200) were used as an alternative. The cancer-related plants and compound structures for Malay TM were derived from a commercial database, Natural Product Discovery System (NADI) (201). The cancer-related plants and their structures were compiled by using keywords "tumour" and "cancer" in the search engine of the database.

In total, the numbers of cancer-related plants of TCM, Ayurveda, and Malay TM were 47, 34, and 28 respectively. Due to identical plants being observed in the TMs, the set of compounds of those particular plants were revised so that the set of compounds of each plant was consistent. This was done by combining the set of compounds of the same plants followed by removing duplicate compounds. The final numbers of compounds were 2,292 (TCM), 1,583 (Ayurveda) and 1,127 (Malay TM) (**Table 3.1**).

Table 3.1: The dataset of TCM, Ayurveda and Malay TM. It can be seen that the distribution of the compounds is variable. The order of the plants follows the order in the phylogenetic tree from right to left.

Plant	TCM	Ayurveda	Malay TM	No of compounds
<i>Pteris multifida</i>	16			16
<i>Drynaria fortunei</i>	50			50
<i>Cedrus deodara</i>		6		6
<i>Schisandra chinensis</i>	66			66
<i>Cinnamomum aromaticum</i>	72			72
<i>Magnolia officinalis</i>	108			108
<i>Annona squamosa</i>		41		41
<i>Houttuynia cordata</i>	27			27
<i>Piper betle</i>		18		18
<i>Piper nigrum</i>			35	35
<i>Aloe vera</i>		139	139	278
<i>Allium sativum</i>		163	163	326
<i>Imperata cylindrica</i>	18			18
<i>Cymbopogon citratus</i>		78	78	156
<i>Curcuma longa</i>	89	89	89	267
<i>Kaempferia galanga</i>			2	2
<i>Alpinia officinarum</i>	99			99
<i>Zingiber montanum</i>			9	9
<i>Zingiber officinale</i>	76	76		152
<i>Aconitum kusnezoffii</i>	9			9
<i>Pulsatilla chinensis</i>	27			27
<i>Nigella sativa</i>		20		20
<i>Paeonia suffruticosa</i>	28			28
<i>Portulaca oleracea</i>	40			40
<i>Persicaria orientalis</i>	45			45
<i>Polygonum aviculare</i>	22			22
<i>Polygonum cuspidatum</i>	45			45
<i>Vitis vinifera</i>		266		266
<i>Terminalia arjuna</i>		42		42
<i>Lawsonia inermis</i>			10	10
<i>Capparis spinosa</i>		39		39
<i>Moringa oleifera</i>		35	35	70
<i>Raphanus sativus</i>		57		57
<i>Isatis tinctoria</i>	103			103
<i>Anacardium occidentale</i>		19		19
<i>Boswellia serrata</i>		22		22
<i>Citrus aurantium</i>	42			42
<i>Zanthoxylum nitidum</i>	41			41
<i>Melia azedarach</i>		46		46
<i>Azadirachta indica</i>		30		30
<i>Eurycoma longifolia</i>			79	79
<i>Brucea javanica</i>	9			9

<i>Samadera indica</i>		16	16
<i>Gynostemma pentaphyllum</i>	47		47
<i>Citrullus lanatus subsp. vulgaris</i>		15	15
<i>Momordica charantia</i>		47	47
<i>Trichosanthes kirilowii</i>	50		50
<i>Phyllanthus niruri</i>		6	12
<i>Ricinus communis</i>		50	50
<i>Euphorbia hirta</i>		15	15
<i>Garcinia atroviridis</i>		6	6
<i>Garcinia mangostana</i>		62	62
<i>Albizia lebbbeck</i>		27	27
<i>Senna tora</i>	49	49	98
<i>Senna alata</i>		4	4
<i>Astragalus complanatus</i>	29		29
<i>Spatholobus suberectus</i>	48		48
<i>Pisum sativum</i>		56	56
<i>Abrus precatorius</i>		32	64
<i>Sophora tonkinensis</i>	37		37
<i>Sophora japonica</i>	17		17
<i>Cornus officinalis</i>	140		140
<i>Diospyros kaki</i>		28	28
<i>Centella asiatica</i>		38	76
<i>Angelica sinensis</i>	90		90
<i>Foeniculum vulgare</i>	38		38
<i>Elephantopus scaber</i>		24	48
<i>Carthamus tinctorius</i>	121		121
<i>Atractylodes lancea</i>	23		23
<i>Saussurea costus</i>		16	16
<i>Gynura procumbens</i>		2	2
<i>Inula britannica</i>	32		32
<i>Eclipta prostrata</i>	28		28
<i>Centipeda minima</i>	29		29
<i>Artemisia argyi</i>	106		106
<i>Chrysanthemum x morifolium</i>	50		50
<i>Chrysanthemum indicum</i>	19		19
<i>Heliotropium indicum</i>		6	6
<i>Cuscuta chinensis</i>	19		19
<i>Datura metel</i>		19	19
<i>Physalis minima</i>		5	5
<i>Withania somnifera</i>		26	26
<i>Calotropis gigantea</i>		16	16
<i>Catharanthus roseus</i>		54	54
<i>Gardenia jasminoides</i>	12		12
<i>Rubia cordifolia</i>		12	12
<i>Oldenlandia diffusa</i>	22		22
<i>Forsythia suspensa</i>	86		86
<i>Andrographis paniculata</i>		21	42
<i>Bacopa monnieri</i>		9	9

<i>Picrorhiza kurrooa</i>		11		11
<i>Scutellaria baicalensis</i>	43			43
<i>Scutellaria barbata</i>	69			69
<i>Lycopus lucidus</i>	11			11
<i>Prunella vulgaris</i>	45		45	90
<i>Ocimum tenuiflorum</i>		13		13
<i>Orthosiphon stamineus</i>			34	34
Total	2,292	1,583	1,127	5,002

3.2.2 A phylogenetic tree of anti-cancer plants

The anti-cancer plants were converted into NCBI Taxonomy IDs and standardised scientific names by using the NCBI Taxonomy Database (202). The list of IDs was uploaded to the first web-based tool, Phylogenetic Tree Generator (phyloT) (203) to create an output file of phylogenetic tree in Newick format. The figure of the phylogenetic tree was generated by uploading the Newick file format to the second web-based tool, Interactive Tree Of Life (iTOL) v3.4 (204). A table of source(s) of the TM for each scientific name was created in a separate file to create a bar chart label at the end of each node of the phylogenetic tree.

3.2.3 Structural similarities between TM compounds and anti-cancer drugs

Both TM compounds and 151 anti-cancer drugs of DrugBank (30), in SMILES (simplified molecular-input line-entry system) format, were characterised as 2,048 bit of ECFP_4 circular Morgan-type fingerprints (Figure 3.2) using RDkit (205). The generation of ECFP_4 fingerprints can be described in three steps (48). First, each heavy atom of a molecule is assigned with an identifier. Second, the features of each atom are described through two iterations that equal to four bonds diameter from the identifier. At iteration zero, the information only represents the initial atom identifier (atom 1). At iteration one, the information describes the immediate neighbours (atom 2) of atom 1. At iteration two, the information describes the immediate neighbours of atom1 and the immediate neighbours of atoms 2. The iteration step is repeated for all identifiers in a molecule. Third, duplicates of the same feature are removed by keeping only one representative of that particular feature. As a result, the fingerprint of a molecule contains sub-structural information from all parts of the molecule.

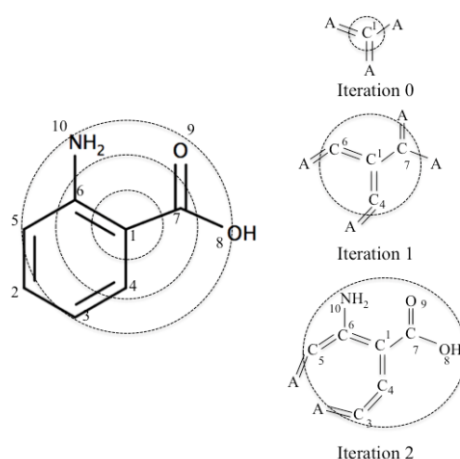


Figure 3.2: The figure illustrates an example of how an ECFP4 fingerprint is generated for C1 of a molecule. Each iteration describes larger and larger circular neighbourhoods of C1 atom. After two iterations, the description of C1 atom has grown further to include amide group, carboxylic group and as much of the aromatic ring.

After generating the ECFP₄ fingerprints, the pairwise similarity combinations were calculated using Tc. The conversion was done in a workflow built in KNIME v.2.10.4 (206). Only Tc values between TM compounds and anti-cancer drugs were extracted for analysis.

3.2.4 Comparison of scaffold diversity

The compound structures of the previously generated SMILES format were uploaded to DataWarrior v4.2.2 (207). The frameworks of the compounds were decomposed into Murcko scaffolds (208). Each type of scaffold was given a unique identifier for comparison purposes. To measure scaffold diversity, for each TM dataset, a ratio between the number of unique scaffolds and the total number of compounds was calculated. The ratios of the Order/Family/Genus to quantity unique scaffolds were measured to determine whether the scaffold diversity was influenced by the taxon. The distribution of scaffold diversity of each TM dataset was measured using cyclic systems retrieval (CSR) curves (209). This was done by plotting two types of fractions, fraction of scaffolds and fraction of database, from the most common scaffolds to the least common scaffold. The fraction of scaffolds was defined as the ratios of compounds to scaffold types, while the fraction of database was defined as the ratios of compound of the scaffold types to the total number of compounds. The diversity of a TM dataset was determined by the fraction of scaffolds accounting for 50% of the TM compounds, which is also known as F₅₀ (209).

3.2.5 Compounds pre-processing

The compounds in SMILES format were uploaded to Standardizer (210) to standardise the compounds prior to *in silico* target prediction. Six pre-selected actions were specified in the configuration file; 'Remove Fragment', 'Remove Explicit Hydrogens', 'Neutralize', 'Clean 2D', 'Mesomerize' and 'Tautomerize'. The standardised TM compounds were saved in SMILES format.

3.2.6 Target prediction

The standardised compounds were run on an *in silico* target prediction algorithm that was modeled using the Random Forest (211). Briefly, the model contained a training set of active compounds extracted from ChEMBL v.21.0 (28) and inactive compounds extracted from the PubChem Compound and PubChem BioAssay databases (212). In total, the model contained 5,888,615 ligand-target pairs of actives and inactive associate with 1,651 targets. The active training set were active compounds with reported activities (K_i/K_d/IC₅₀/EC₅₀) of lower than 10 μM with a CONFIDENCE_SCORE of 5 or greater for 'binding' or 'functional' human

protein assays (28). The inactive training set were inactive compounds of assays submitted by researchers to PubChem Compound and PubChem BioAssay databases (28). The descriptor of a molecule was generated based on ECFP_4 circular Morgan-type fingerprints (48) with a 2,048 bit length, hence stereochemistry of compounds was excluded. The algorithm calculates the probability of an orphan compound belonging to a target class by considering both chemical features of active and inactive compounds against target classes, giving a more holistic perspective of chemical features, which contribute towards or against bioactivity.

To predict targets of TM compounds, the Random Forest (RF) classifier can be briefly described as follows. First, a Scikit-learn (Version 17) RF classifier of 100 trees, with the number of features and max depth set to ‘*auto*’ and the ‘*class_weight*’ set to ‘*balanced*’, was trained using the binary matrix of the active and inactive compound-target fingerprints on a per-target basis (213). Models generate a probability representing the number of trees in the forest that predict a target as active, represented as $p(\text{activity})$. Platt scaling (214), a calibration method in machine learning, is performed to convert the classification from the random forest into the probabilities of the prediction being correct, which is known as True Positive Rate (TPR).

3.2.7 The TPR of predicted targets

The TPR value was selected by assessing the distribution of fraction of targets predicted for each TPR interval of 0.1 between 0 and 1. A curve of the fraction of targets predicted TPR intervals was evaluated to define a TPR threshold of each TM (**Figure 3.3**). In general, the fraction of targets predicted decreases as the threshold value increases. A threshold value of 0.7 was selected for all TM datasets because at least 60% of the targets in the model were predicted. The selected threshold value is a reasonable trade-off between false-negative predictions and number of predicted targets compares to larger threshold values. By using the selected TPR threshold, the target predictions algorithm was performed, which produced a binary matrix of predicted targets.

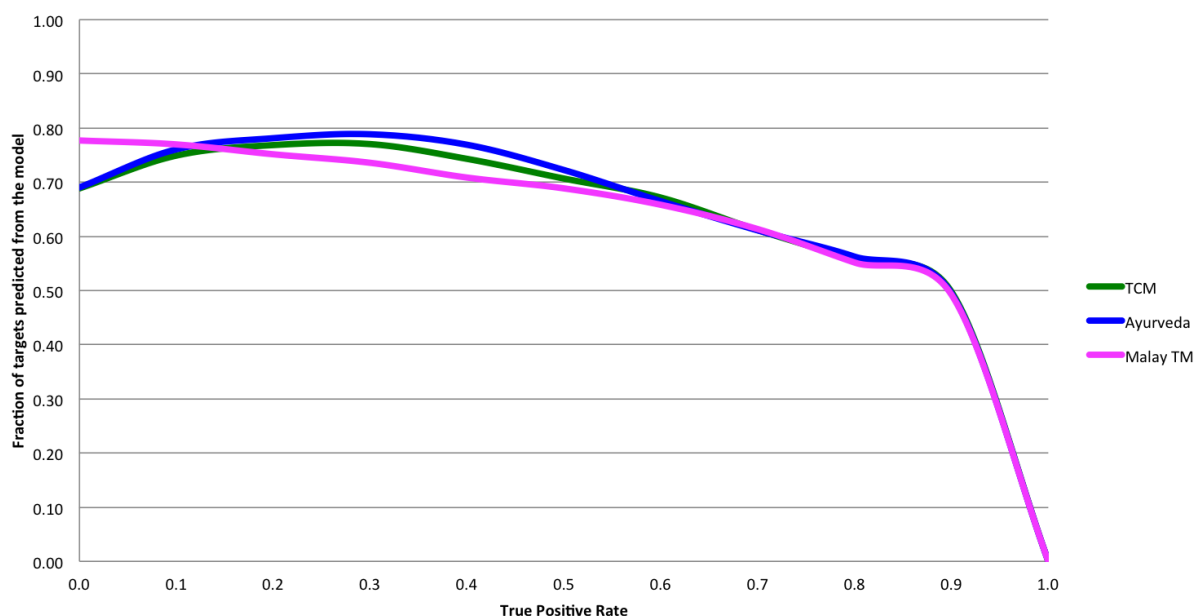


Figure 3.3: The curves of the fraction of targets predicted decreases when TPR values increases. A TPR threshold value of 0.7 was selected in this analysis because more than 60% of the targets were predicted.

3.2.8 Annotating predicted targets as cancer-related targets

The UniProt IDs of the predicted targets were mapped to their gene names in the *Retrieve/ID mapping* tab in UniProt (116). We found that three UniProt IDs, P47929, Q13748, and Q16637 were annotated with two gene names, while P62158 was annotated with three gene names. Only primary gene names defined by Uniprot (116) were used in the analysis. Two pairs of UniProt IDs, A2TJX0, Q13887 and P0DMS8, P33765 were synonyms. All the gene names were uploaded to Target Validation Platform (215) to annotate the gene names with cancer. By using the keyword ‘cancer’ in the search tab of the Target Validation Platform (215), a total of 28,019 genes associated with cancer was listed. The list was compared to the gene names of predicted targets.

3.2.9 Heatmap with hierarchical clustering based on bioactivity space and target class

Two heatmaps were generated in this study. The first heatmap compared the bioactivity space similarities of the TM plants, while the second heatmap compared the targets classes of TMs. Before constructing the heatmaps, the frequencies of compounds of per predicted target and the frequencies of predicted target per target class were normalised. The dissimilarity distances between the normalised frequencies of TM plants/target classes were calculated using ‘Euclidean’ method of the “dist” function. Clustering was performed using the Ward method of the “hclust” function, where two TM plants/target classes were merged if the sum of squared Euclidean distance was minimal. The heatmaps were generated using “heatmap” function. All the aforementioned functions were available in the stats package of R (114).

3.3 Results and Discussions

3.3.1 Phylogenetic tree of anti-cancer plants

In this study, a total of 97 anti-cancer plants were used to construct a phylogenetic tree (**Figure 3.4**) and each plant was annotated with its source of TM (TCM/Ayurveda/Malay TM). Altogether, nine identical plants were found in Ayurveda and Malay TM, two identical plants were observed between TCM and Malay TM, one identical plant was identified in TCM and Ayurveda and one plant was present in all three TMs. The distribution of the labelled TM plants is observed to be scattered along the phylogenetic tree and is not clustered according to the type of TM indicating that, in many cases, the plants from different TMs share common ancestors. In other words, TMs are, according to this analysis, intrinsically related to each other. Each terminal taxon is marked with filled circles that represent taxonomic ranks. For each plant, the rings around the filled circles represent superkingdom, kingdom, phylum, order, family and genus. These taxonomic ranks were consistently annotated. The remaining ranks that were annotated varied from class, subclass, subfamily, tribe, subtribe, subgenus, and subspecies.

Figure 3.4 shows that all plants are rooted to phylum, Streptophyta. The TM plants were annotated with 27 orders, 48 families, and 90 genera, indicating that the most immediate evolutionary relationship of the plants is found at genus level. There are seven clades that share the same genus, *Piper betle* (Ayurveda) and *Piper nigrum* (Malay TM), *Zingiber montanum* (Malay TM) and *Zingiber officinale* (TCM and Ayurveda), *Polygonum aviculare* (TCM) and *Polygonum cuspidatum* (TCM), *Garcinia atrovirdis* (Malay TM) and *Garcinia mangostana* (Malay TM), *Senna tora* (TCM and Malay TM) and *Senna alata* (Malay TM), *Sophora tonkinensis* (TCM) and *Sophora japonica* (TCM), *Chrysanthemum x morifolium* (TCM) and *Chrysanthemum indicum* (TCM), and *Scutellaria baicalensis* (TCM) and *Scutellaria barbata* (TCM). For at least three pairs plants, the findings indicate that plants of different TMs with the same medicinal usage are evolutionary related. A similar observation was demonstrated by Saslis-Lagoudakis et. al (216). The study of the genus *Pterocarpus* that is geographically distributed in Neotropics, Tropical Africa and Indomalaya indicates the same ethnomedicinal uses of *Pterocarpus* plants, which have been largely used for gastrointestinal and skin illnesses. At a molecular level, the similar medicinal properties of plants of shared phylogeny was investigated by Lopez et. al (217). The screening of alkaloid extracts of 26 wild plants belonging to the genus *Narcissus* L., for acetyl cholinesterase (AChE) inhibitory activity shows that the lowest IC₅₀ values are expressed by those extracts containing galanthamine.

In general, the TM plants share different genera, which can be explained by convergent evolution (218). Independent evolution of the same trait is a result of unrelated plants in the phylogeny to adapt to similar environmental challenges. Hence, similar compounds of different TM plants might be responsible for the observed therapeutic effects. The screening of phytochemical constituents of different plants with overlapped medicinal uses could confirm the distribution of similar bioactive compounds. For example, the investigation of phytochemicals of ten medicinal plants of different families in the Mardan region of Pakistan shows that flavanoids are identified in six medicinal plants, which have been used traditionally for heart disease (219).

Therefore, it was hypothesised that the common MOAs of TM plants for treating cancer would stem from similar compounds, and the phylogenetic distance between the plants could either be small or large. The phylogenetic tree of 97 TM plants was used to determine whether proximity in the phylogenetic tree resembled the similarity of chemical as well as cancer-related bioactivity space of the compounds of those plants.

3.3.2 Comparison of chemical space between TM compounds and anti-cancer drugs

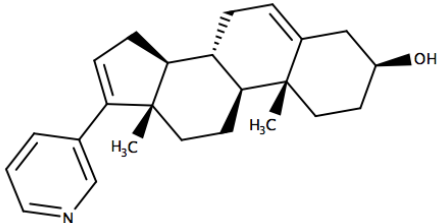
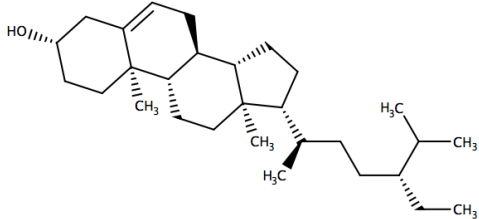
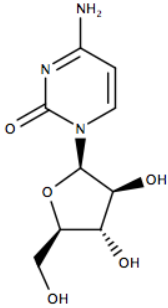
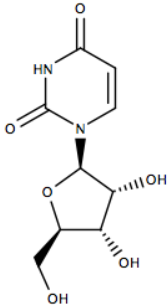
We next analysed the chemical space similarities between TM compounds and 151 anti-cancer drugs from DrugBank (30). In this study, the Tc value of 0.3 was used to determine whether a TM compound is structurally similar to an anti-cancer drug, considering the previous report on the activity–relevant similarity of active compounds (220). Only 204 TM compounds (including identical compounds of different plants) were structurally similar to 28 anti-cancer drugs. The list of 28 anti-cancer drugs and the similar TM compounds can be found in **Table A3.1** (Appendices).

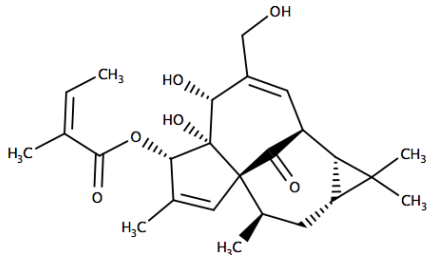
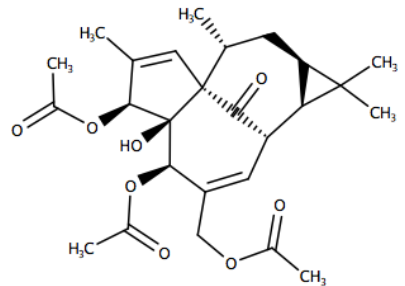
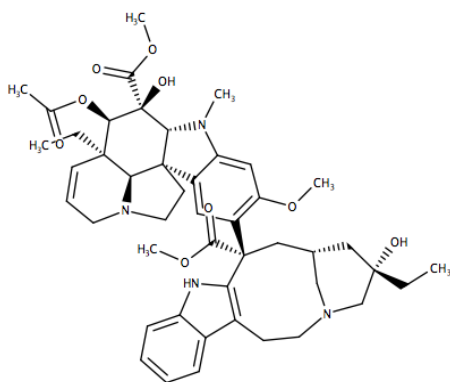
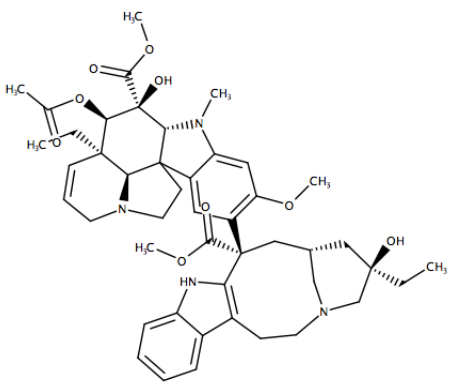
The observation of **Table A3.1** shows that the anti-cancer drugs can be classified into derived from natural products or resembling metabolites. Among the anti-cancer drugs in the first group are Abiraterone (DB05812), Cytarabine (DB00987), Ingenol Mebutate (DB05013), Vinblastine (DB00570), Vincristine (DB00541), Vindesine (DB00309) and Vinorelbine (DB00361) are derived from plants (**Table 3.2**), while the remaining anti-cancer drugs are derived from bacteria (221–224). Abiraterone (DB05812) is naturally derived from ketoconazole, an anti-fungal agent, which has caused serious side effects when the compound is given to prostate cancer patients (225, 226). Abiraterone has been used for castration resistant prostate cancer in men, in which the drug modulates CYP17, a catalysis enzyme for the production of androgen, to inhibit the proliferation of prostate cancer cells (227). It was discovered that beta-sitosterol, which is structurally similar to Abiraterone (DB05812) (Tc = 0.331), is widely distributed in 40 different plants (**Table A3.1**). Structurally, both beta-sitosterol and Abiraterone contain steroid four-fused rings and differ in their side chains. Although beta-sitosterol is not developed into an anti-cancer drug, the compound has been studied extensively and showed to induce apoptosis in cancer cells *in vitro* (228). Cytarabine (DB000987) is inspired by a series of C-nucleoside-derived compounds isolated from the Caribbean sponge, *Cryptotheca crypta*, which is widely used for the treatment of leukaemia and lymphoma (229). Cytarabine shows structural similarities with guanosine (Tc=0.357) and adenosine (Tc=0.333) of *Allium sativum* (Ayurveda and Malay TM), adenosine (Tc=0.333) of *Carthamus tinctorius* (TCM), and uridine (Tc=0.420) of *Isatis tinctorius* (TCM). These TM compounds are also nucleosides. Ingenol Mebutate (DB05013), isolated from *Euphorbia peplus*, has been used for the treatment of skin cancer by inducing cellular necrosis (230). **Table 3.2** shows that Ingenol-triacetate of *Euphorbia hirta* (Ayurveda) is structurally similar to Ingenol mebutate with a Tc value of 0.456. It has been reported that the methanol extract of *Euphorbia hirta* has been evaluated for cytotoxic activity with the most effective activity is the inhibition of HEP-2 cell proliferation (231). All four drugs, Vinblastine (DB00570), Vincristine (DB00541), Vindesine (DB00309), and Vinorelbine (DB00361) are derived from

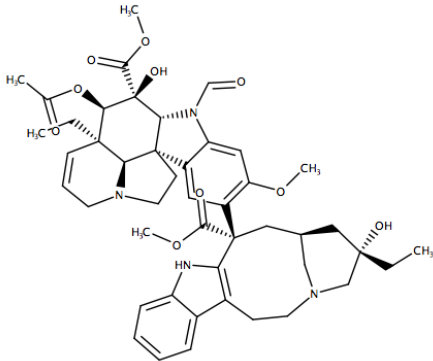
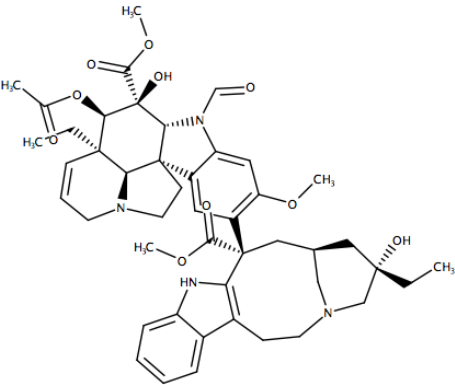
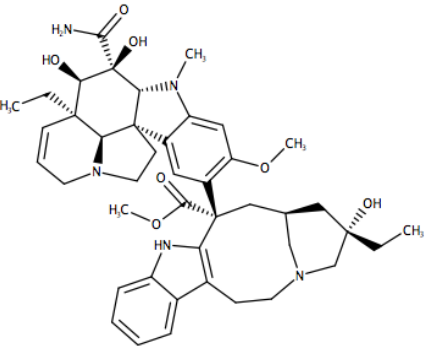
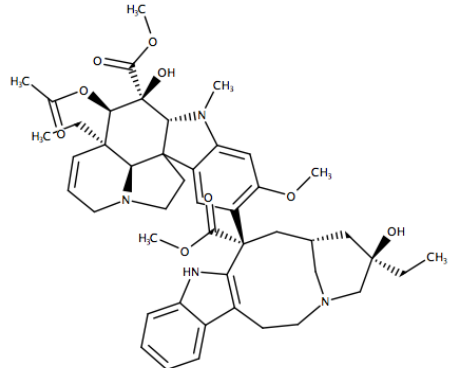
Catharanthus roseus (Malay TM) and two compounds of this plant (leurosidine and leurocristine) showed Tc values of 1.0 to Vinblastine and Vincristine. The MOA of Vinblastine has been reported by binding to tubulin and microtubules, which subsequently depolymerises the microtubules and destroys mitotic spindles to block mitosis of cancer cells (232). Among other TM compounds that are structurally similar to the four drugs is vindoline, which has been reported with cytotoxic activity against the human colorectal carcinoma cell line HCT 116 (233). In the second group, a few of the anti-cancer compounds resemble the naturally occurring compounds in the human body (guanosine, phenylalanine, thymidine, uridine) such as Decitabine (DB01262), Clofarabine (DB00631) and Fluradabine (DB01073). These drugs act by interfering with cellular processes of naturally occurring nucleosides such as nucleic acid synthesis, cell signalling and enzyme metabolism (234). The summary of these seven anti-cancer drugs can be found in **Table 3.2**. Given that only seven anti-cancer drugs, which are derived from plants, are structurally similar to the TM compounds, the findings indicate the possibilities of finding more anti-cancer compounds from TM plants due to many unexplored compounds.

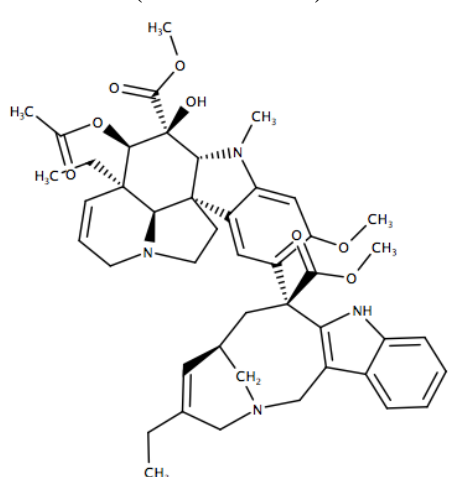
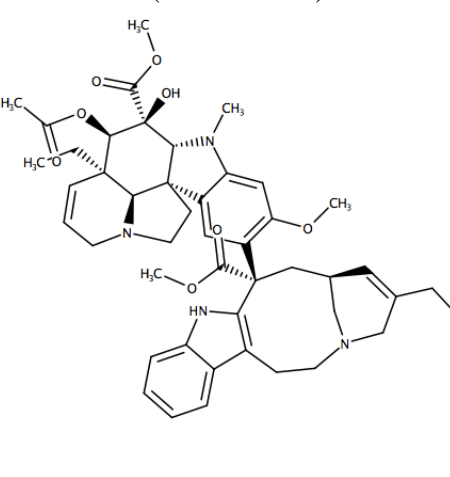
Table 3.2: The structurally similar TM compounds and anti cancer drugs. The complete list of TM compounds that are similar ($T_c \geq 0.3$) to the anti-cancer drugs can be found in Table 1S in Supporting Information. Only the compounds with the highest T_c value are shown here.

*PID : PubChem ID

No	Anti-cancer drugs	DrugBank ID	TM compounds	Plants	Source of TM	Tc value
1	<p>Abiraterone (PID 132971)</p> 	DB05812	<p>Bassicaterol (PID 5281327)</p> 	<p><i>Houttuynia cordata</i> <i>Moringa oleifera</i></p>	TCM Ayurveda, Malay TM	0.339
2	<p>Cytarabine (PID 6253)</p> 	DB00987	<p>Uridine (PID 6029)</p> 	<i>Isatis tinctoria</i>	TCM	0.420

3	<p>Ingenol Mebutate (PID 6918670)</p> 	DB05013	<p>Ingenol-Triacetate (PID 65377)</p> 	<i>Euphorbia hirta</i>	Ayurveda	0.456
4	<p>Vinblastine (PID 13342)</p> 	DB00570	<p>Leurosidine (Vinrosidine) (PID 161115)</p> 	<i>Catharanthus roseus</i>	Malay TM	1.00

5	<p>Vincristine (PID 5978)</p>  <p>The chemical structure of Vincristine is a complex pentacyclic alkaloid. It features a tetracyclic core with a phenanthrene-like system. Key functional groups include a methyl ester, a hydroxyl group, a methoxy group, and a dimethylamino group. The structure is shown with stereochemistry indicated by wedges and dashes.</p>	DB00541	<p>Leurocristine (PID 5978)</p>  <p>The chemical structure of Leurocristine is very similar to Vincristine, but it lacks the dimethylamino group at the C-4 position, instead having a hydroxyl group. The rest of the structure, including the methyl ester, methoxy groups, and the complex ring system, is identical to Vincristine.</p>	<i>Catharanthus roseus</i>	Malay TM	1.00
6	<p>Vindesine (PID 40839)</p>  <p>The chemical structure of Vindesine is a complex pentacyclic alkaloid. It features a tetracyclic core with a phenanthrene-like system. Key functional groups include a methyl ester, a hydroxyl group, a methoxy group, and a dimethylamino group. The structure is shown with stereochemistry indicated by wedges and dashes.</p>	DB00309	<p>Leurosidine (Vinrosidine) (PID 161115)</p>  <p>The chemical structure of Leurosidine (Vinrosidine) is very similar to Vindesine, but it lacks the dimethylamino group at the C-4 position, instead having a hydroxyl group. The rest of the structure, including the methyl ester, methoxy groups, and the complex ring system, is identical to Vindesine.</p>	<i>Catharanthus roseus</i>	Malay TM	0.733

7	<p>Vinorelbine (PID 5311497)</p>  <p>The chemical structure of Vinorelbine is a complex pentacyclic alkaloid. It features a tetracyclic core with a piperidine ring fused to a pyridine ring, which is further fused to a cyclohexene ring. A quaternary carbon atom is bonded to a methyl group, a hydroxyl group, a methoxycarbonyl group, and a side chain. The side chain consists of a methoxy group, a methylene group, and a piperidine ring fused to a benzene ring. The structure is shown with stereochemistry indicated by wedges and dashes.</p>	DB00361	<p>3',4'-Anhydrovinblastine (PID 443324)</p>  <p>The chemical structure of 3',4'-Anhydrovinblastine is a complex pentacyclic alkaloid. It features a tetracyclic core with a piperidine ring fused to a pyridine ring, which is further fused to a cyclohexene ring. A quaternary carbon atom is bonded to a methyl group, a hydroxyl group, a methoxycarbonyl group, and a side chain. The side chain consists of a methoxy group, a methylene group, and a piperidine ring fused to a benzene ring. The structure is shown with stereochemistry indicated by wedges and dashes.</p>	<i>Catharanthus roseus</i>	Malay TM	0.756
---	---	---------	---	----------------------------	----------	-------

3.3.3 Molecular scaffold analysis

The decomposition of the compounds' frameworks based on Murcko scaffolds (208) produced 776 unique scaffolds, which represented 76.63% of the compounds, while the remaining compounds were identified as macrocyclic and acyclic compounds (23.37%). From 776 scaffolds, the highest common scaffold, 14.18%, was between Ayurveda and Malay TM, owing to the same plants found in both TMs. The three TMs shared 13.40% common scaffolds. The common scaffolds of TCM-Ayurveda and TCM-Malay TM was 3.87% and 3.22% respectively. Nevertheless, each TM comprised a set of distinct scaffolds, which was 29.12%, 18.30%, and 17.91% for TCM, Ayurveda, and Malay TM respectively. The higher percentages of distinct scaffolds than common scaffolds imply that the TMs were diverse from each other.

To compare the diversity of the three TM datasets, various ratio metrics (N/M , N_{sing}/N , and N_{sing}/M) (190) were calculated in **Table 3.3**. In general, the higher the ratio, the more diverse the dataset is. From this analysis Malay TM is highly diverse followed by Ayurveda and TCM. In addition to ratio metrics, cyclic systems retrieval (CSR) curves, which have been used previously to assess the distribution of compound in cyclic systems, were generated (**Figure 3.5**) (190, 209). The curves also show that Malay TM is the most diverse dataset followed by Ayurveda and TCM, with the F_{50} values of TCM, Ayurveda, and Malay TM are 0.018, 0.028, and 0.052. Despite having the largest set of compounds, TCM has the least diverse scaffolds, where 50% of the TCM compounds were decomposed into six types of scaffolds. At one-half of the datasets, Ayurveda and Malay TM, compounds were decomposed into eleven and twenty types of scaffolds respectively.

The low scaffold diversity of the TCM dataset is not influenced by the phylogeny where the ratio of scaffold to order and scaffold to genus are the highest compared to those of Ayurveda and Malay TM. Many similar compounds of TCM belong to unrelated plants. On the contrary, the high scaffold diversity of the Malay TM data set is influenced by the phylogeny where the ratio of scaffold to order, scaffold to family and scaffold to genus were the lowest. The results can be seen in the phylogenetic tree, where, in many cases, Malay TM plants are closely clustered. In brief, the findings demonstrate that TCM agrees with the convergent evolution process whereas the diverse scaffolds observed in Malay TM can be explained by divergent evolution (186).

Table 3.3: Results of scaffolds diversity analysis.

Database	M	N	O	F	G	N/M	Nsing	Nsing/N	Nsing/M	O/N	F/N	G/N
TCM	2292	391	20	24	43	0.17	213	0.54	0.09	0.051	0.061	0.110
Ayurveda	1583	392	19	28	36	0.25	236	0.60	0.15	0.048	0.071	0.092
Malay TM	1127	383	16	19	26	0.34	252	0.66	0.22	0.042	0.050	0.068

M=No of compound in database, **N**=No of unique scaffolds, **O**=No of Order taxon, **F**=No of Family taxon, **G**=No of Genus taxon, **Nsing**=No of singleton scaffolds

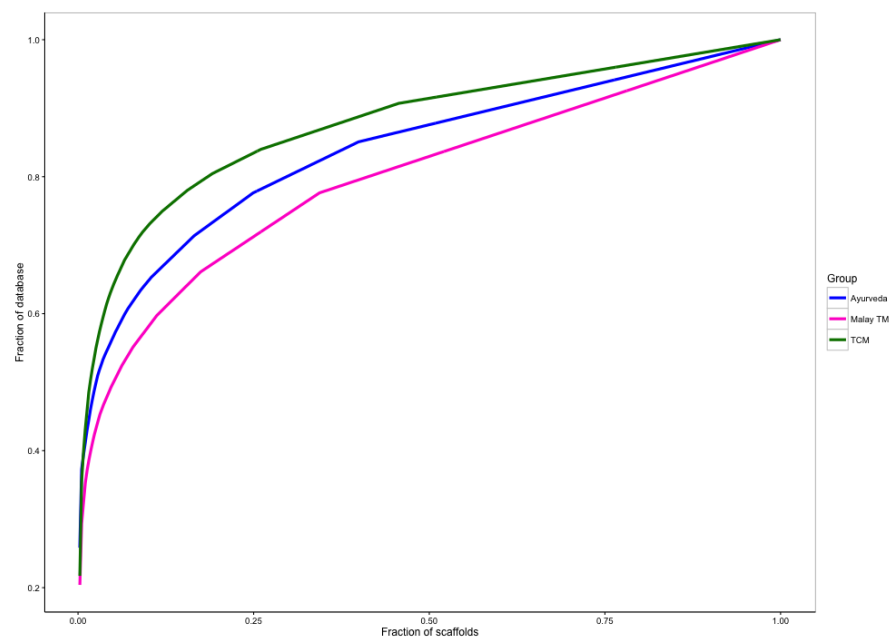


Figure 3.5: The cyclic systems retrieval (CSR) curves indicate Malay TM has higher scaffold diversity than Ayurveda and TCM. At one-half of the dataset, the compounds of Malay TM were decomposed into 20 scaffold types compared to Ayurveda (eleven scaffold types) and TCM (seven scaffold types).

To further analyse the prevalent scaffolds of TMs, we looked at the top ten most frequent scaffolds (**Figure 3.6**). Overall, six scaffolds were frequently present in all TMs, two scaffolds were frequently present in Ayurveda and Malay TM and one scaffold was frequently present in Ayurveda and TCM. No common scaffold was observed between TCM and Malay TM. Each TM showed three (TCM), two (Malay TM), one (Ayurveda) different scaffolds from each other. It was found that the benzene scaffold was the highest frequent scaffold identified across all TMs, which also has been commonly identified in other natural product databases (190). A few of these scaffolds are already explored scaffolds in drug discovery such as anthraquinone (235), coumarin (236), cyclohexene (237), flavone and isoflavone (238), sterol (239), pentacyclic triterpenes (240) and xanthone (241) with various or focused therapeutic activities. Regarding the cancer-related scaffolds anthraquinone, flavone, sterol and pentacyclic triterpenes are four frequent scaffolds in all TMs.

Anthraquinone is a tricyclic planar ring system with various therapeutic properties (235). The reported MOA of anthraquinone derivatives is primarily by intercalating its aromatic structure with DNA between base pairs to possibly inhibit DNA replication in cancer cells (235). The substitution patterns of the anthraquinone scaffold also contribute to the anti-cancer properties of the compounds (235). Flavone contains three-ring skeletons, and the MOAs of flavone derivatives are diverse. The functional groups of the scaffold makes it amenable to selective modulators for different targets (238). Phytosterol, which contains a tetracyclic lipid component, is implicated in cholesterol metabolism (239). One of the MOAs is by lowering blood cholesterol by incorporating less cholesterol in the lipid rafts of cancer cells, thus, promoting apoptosis through signal transduction (239). Pentacyclic triterpene is one of the widely spread secondary metabolites in plants with diverse biological properties. The MOAs of betulinic, ursolic, oleanolic and maslinic acids, a few of the representative compounds of the group, are implicated in cell apoptosis by activating either intrinsic (mitochondrial) or extrinsic (death receptor) apoptotic pathways (240). Xanthone, which is more frequently present in Malay TM, contains a three-membered heterocyclic ring. Similar to the anthraquinone, xanthone derivatives are also DNA-intercalating modulators (241). The stereochemistry of xanthonic derivatives is described to influence their MOAs (242). Coumarin, which is more frequently present in TCM, consists of a benzene ring joined with a pyrone ring. The MOAs of coumarin derivatives are diverse through the diverse substitution of the coumarin nucleus (236). Although cyclohexene is among the most frequently present scaffolds in all TMs, this scaffold is more known as an anti-influenza drug, Oseltamivir (243) than that anti-cancer scaffold.

To summarise, the most frequent scaffolds in the TMs, in many cases, are widely explored for their anti-cancer properties, where the MOAs can either stem from the core structure or functional group substitutions. The compounds are primarily reported to induce apoptosis of cancer cells. The remaining scaffolds in the top ten frequent scaffolds could be worth exploring for their anti-cancer activities, as they are found to abundance in anti-cancer plants.




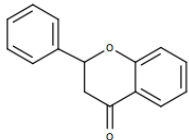

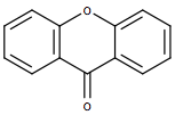
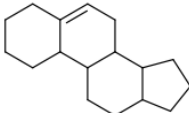
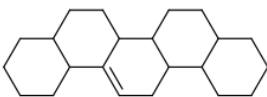
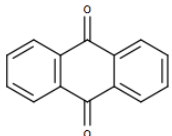
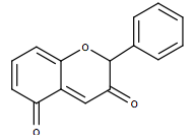

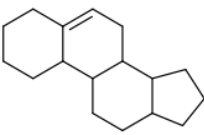

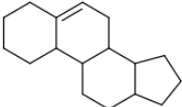
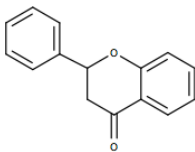

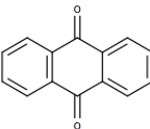

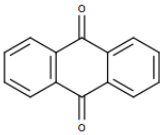
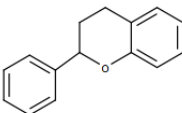
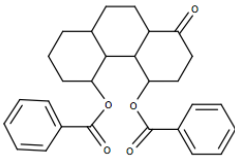
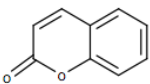
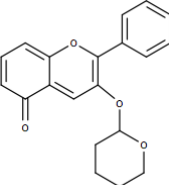

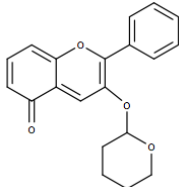



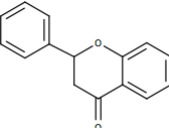

TCM		Ayurveda		Malay TM	
	13.4%		11.43%		8.96%
	4.28%		1.96%		3.02%
	3.27%		1.83%		2.93%
	2.79%		1.83%		1.86%
	2.71%		1.58%		1.51%
	1.88%		1.58%		1.33%
	1.61%		1.39%		1.24%
	1.48%		1.26%		1.15%
	1.48%		1.14%		0.98%
	1.27%		1.14%		0.98%

Figure 3.6: The top ten most frequent scaffolds of TMs. Six scaffolds are frequent in all TMs, two scaffolds are frequent in Ayurveda and Malay TM and one scaffold is frequent in Ayurveda and TCM. No common scaffold is observed between TCM and Malay TM only. Each TM shows three (TCM), two (Malay TM), one (Ayurveda) different scaffolds that are more frequent in that TM than the remaining two TMs.

3.3.4 Comparison of bioactivity space of the plants

We next performed the comparison of the bioactivity space. Out of 1,651 targets in the *in silico* target prediction model, 1,163 were predicted. All the predicted targets were annotated with cancers when compared to the list of associated cancer targets from a Target Validation Platform (215). The normalised frequency of predicted targets was used to generate a heatmap in **Figure 3.7**, which compares the bioactivity space of TM plants.

Overall, the majority of TM plants show similarities of their bioactivity space but they are not phylogenetically related, except in a few cases. The similarities of the bioactivity space between TM plants with relation to their phylogeny can be observed in three clusters, I, II and III, which are labelled on the left dendrogram of the heatmap. The main difference of the three clusters is that the plants of Cluster I contain a higher number of compounds than plants in Cluster II and III, hence more similar predicted targets are observed between plants in Cluster I. Despite the sparse distribution of compounds in TM plants, the heatmap shows that the similar bioactivity space between eight pairs of plants could be linked to their phylogeny either at the order, family or genus taxonomic rank in all three clusters. The top three frequently annotated protein families of shared predicted targets are selected to describe the observation. In cluster I, two pairs of plants are shown with similar bioactivity space. The first pair, *Alpinia officinarum* (TCM) and *Zingiber officinale* (Ayurveda and Malay TM) share the same family (Zingiberaceae) and order (Zingiberales). The second pair, *Alium sativum* (Ayurveda and Malay TM) and *Aloe vera* (Ayurveda and Malay TM) are members of the same order, Asparagales. The percentages of identical predicted targets are 67.54% and 77.53% in both pairs respectively. The compounds of those four plants were predicted to modulate targets of all protein families, and the top five frequent target classes are kinase, oxidoreductase, ion channel, GPCR and hydrolase.

In the Cluster II, three pairs of plants are identified to share similar bioactivity space, namely, *Spatholobus suberectus* (TCM) and *Sophora tonkinensis* (TCM), and *Eclipta prostrata* (TCM) and *Chrysanthemum x morifolium* (TCM) are members of the family, Fabaceae and Asteraceae, and the order, Fabales and Asterales, respectively. The third pair of the plants, *Scutellaria barbata* (TCM) and *Scutellaria baicalensis* (TCM), is phylogenetically related at the genus level, Scutellaria. The percentages of overlapped bioactivity space for each pair of the plants are 73.29%, 72.05% and 70.56% respectively. The family Fabaceae and Asteraceae share bioactivity spaces that are frequently annotated to kinase, oxidoreductase, ion channel, GPCR and transferase. The pair of genus Scutellaria plants was also frequently annotated with the same target classes as previous pairs except the fifth target class was hydrolase.

In Cluster III, three pairs of plants are phylogenetically related at the family and order level. The first pair, *Ocimum tenuiflorum* (Ayurveda) and *Lycopus lucidus* (TCM) share the same family (Lamiaceae) and order (Lamiales). The top five target classes of the overlapped predicted targets between these plants are kinase, hydrolase, oxireductase, nuclear hormone receptor (NHR) and GPCR. The final two pairs of plants of this cluster are only phylogenetically related at the order level. The first pair, *Phyllanthus niruri* (Ayurveda and Malay TM) and *Garcinia atroviridis* (Malay TM) are members of the order Malpighiales. The second pair, *Picrorhiza kurrooa* (Ayurveda) and *Orthosiphon stamineus* (Malay TM) belongs to the order, Lamiales. Both plants of the order, Malpighiales and Lamiales were predicted to modulate targets that are frequently annotated to oxireductase, kinase, transferase, lyase, and hydrolase. The percentages of identical predicted targets are 66.10%, 41.20% and 42.90% for the first, second and third pair.

Out of eight pairs of plants with similar bioactivity space and phylogenetically related, only one pair of plants, *Scutellaria barbata* (TCM) and *Scutellaria baicalensis* (TCM) is related at genus level. The genus *Scutellaria* is widely studied for various medicinal properties, and phenols and terpenes have been demonstrated with anti-cancer properties (244). Five pairs of plants belong to the same source of TM except for *Alpinia officinarum* (TCM) and *Zingiber officinale* (Ayurveda and Malay TM), *Ocimum tenuiflorum* (Ayurveda) and *Lycopus lucidus* (TCM) and *Picrorhiza kurrooa* (Ayurveda) and *Orthosiphon stamineus* (Malay TM). Out of the five pairs of plants, three of them, from Cluster II, belong to TCM, which could contribute to the low scaffold diversity of the TCM dataset found in the previous section. The annotation of the target classes demonstrates that kinase and oxireductase are the two most frequent annotated target classes in all eight pairs of plants. In conclusion, generally, TM plants agree with the convergent evolution process, where these plants share common MOAs, except in a few cases, where the common MOA can be described from their relations in the phylogeny.

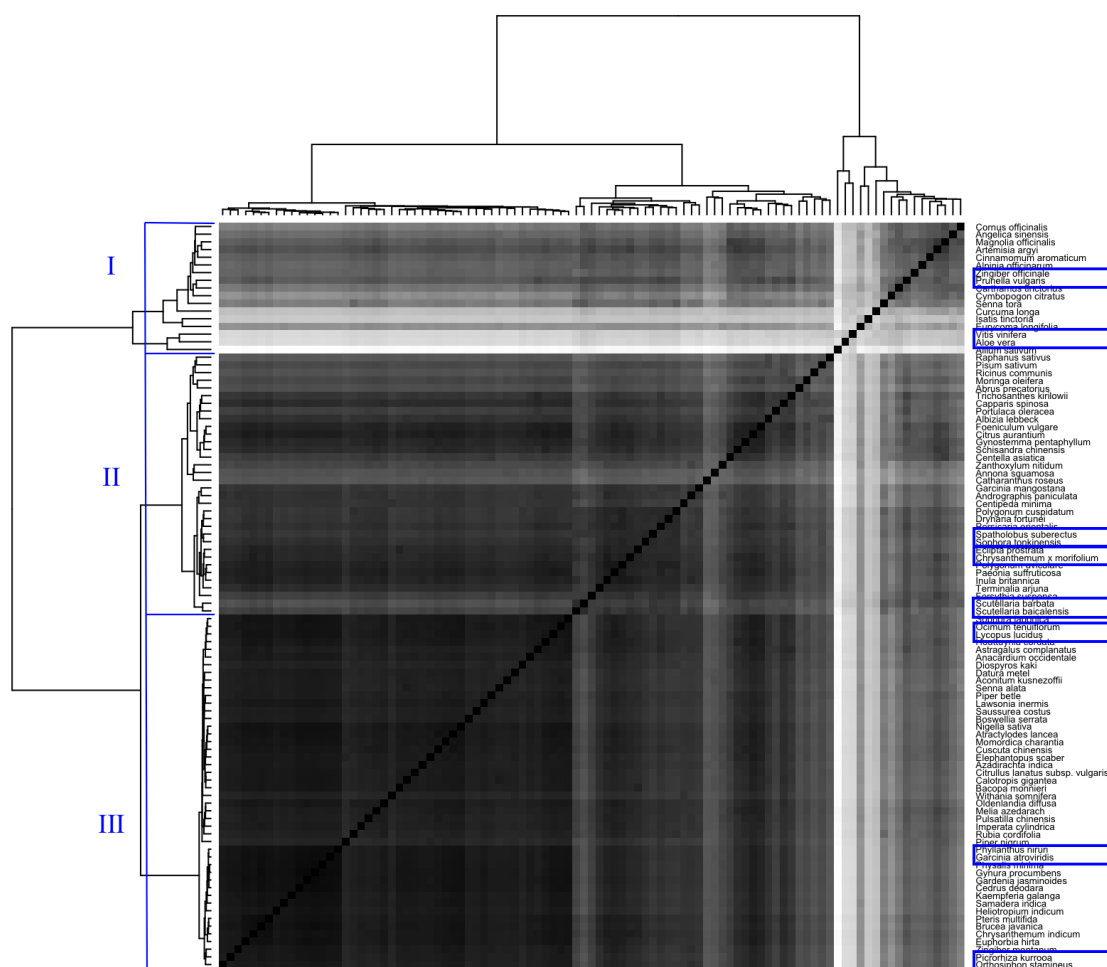


Figure 3.7: The heatmap compares the bioactivity space of the TM plants based on the normalised numbers of compounds modulating the predicted targets. In general, a large number of TM plants share similar bioactivity space although they are not phylogenetically related. Only eight pairs of plants are shown to be phylogenetically related when compared to **Figure 3.2**. These observations can be described by dividing the heatmap into three clusters, Cluster I, II, and III, which for each cluster, at least two pairs of plants (blue box) are phylogenetically related and share similar bioactivity space. The evolutionary relationship of these plants can either be from the genus, family or order taxonomic rank.

3.3.5 Comparison of targets classes of traditional medicines

In this part of study, the aim was to determine the significant target classes of TMs by classifying the predicted targets to their respective target classes. The 1,163 predicted targets were annotated to 14 target classes and one type of target class, 'Other', of which the target class of the remaining targets were not classified. The heatmap in **Figure 3.8** compares the 14 target classes between the TMs. The more saturated colour across the TMs the more significant the target class. The number of the target classes in the TMs was normalised due to the differences in numbers of the predicted targets per TM. What can be seen from the heatmap is that all TMs exhibit kinase as the most significant target class. Kinase, which is one of the largest target classes plays many key roles in cellular functions such as signal transduction, cell cycle regulation, cell division, and cell differentiation, which genetic alterations can often lead to cancer (245). Per target class, the results showed that only three target classes, hydrolase, phosphatase and transporter, were more significant in TCM. Four target classes were discovered to be more significant in Ayurveda, which are GPCR, ion channel, oxidoreductase, and protease. Six target classes, isomerase, ligase, lipase, lyase, NHR and transferase, were identified to be more significant in Malay TM. The observations imply that the TMs use could exert diverse MOAs by modulating various target classes, which can explain the holistic approach of the TMs.

In TCM, among the three significant target classes, hydrolase is the most frequently annotated. The frequently predicted targets of hydrolase are histone deacetylases (HDACs) (246). HDAC is involved in the regulation of gene expression of cell proliferation, cell cycle regulation and apoptosis, which mutations of HDAC could lead to cancer (247). It is observed that GPCR is the most frequent annotated target class in Ayurveda. The most frequent predicted targets of GPCR are serotonin receptors (5-hydroxytryptamine). Serotonin receptor is implicated in cancer by acting as a growth factor on several types of tumour (248). In relating to Malay TM, transferase is the most frequent annotated target class, which is frequently annotated to DNA polymerases. DNA polymerase, an enzyme that synthesises DNA, is generally involved in cancer pathogenesis when DNA strand breaks are not prohibited for DNA replication as the function of p53 is loss (249). The difference in target classes suggests that TMs could exhibit differential MOAs for treating cancers.

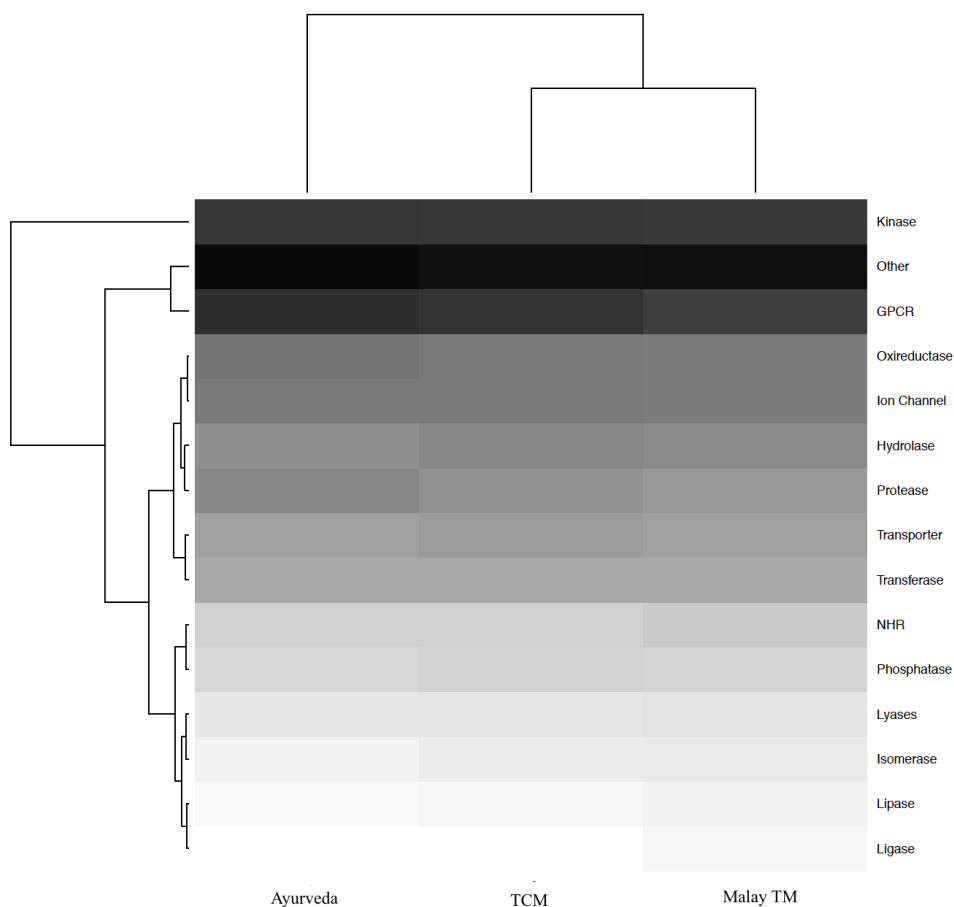


Figure 3.8: The heat map compares the 14 target classes of the TMs based on the normalised numbers of predicted targets. Kinase is observed to be the most significant target class in all TMs. Three, four and six target classes are observed to be more significant in TCM, Ayurveda and Malay TM. The findings indicate that the TMs could exhibit different MOAs for treating cancers.

3.4 Conclusion

TCM, Ayurveda and Malay TM are practiced in different geographical locations but all three TMs have been reported for treating cancers. Given the increasing amount of chemical information available, this study compared both the chemical and bioactivity space that are utilized by such medicines. The phylogenetic tree of 97 TM plants was used to determine whether proximity in the tree resembled the similarity of chemical and bioactivity space. It was found that in these plants were members of 27 orders, 48 families, and 90 genera indicating convergent evolution, where the plants experience similar environmental challenges to result similar anti-cancer properties. In the chemical space analysis, seven anti-cancer drugs, which were derived from plants, were shown to be structurally similar to the TM compounds ($T_c \geq 0.3$). The scaffold diversity analysis demonstrated that Malay TM is the most diverse TM followed by TCM and Ayurveda. The convergent evolution process could explain the low diversity of TCM and the high diversity of Malay TM is a common process of divergent evolution. In the top ten most frequent scaffolds, benzene, anthraquinone, flavone, sterol, pentacyclic triterpene and cyclohexene were the most frequent scaffolds in all TMs. Anthraquinone, flavone, sterol and pentacyclic triterpene have been reported for anti-cancer activities. There are six different scaffolds that are more frequent in one of the TMs than the remaining two TMs, which could be worth exploring for anti-cancer activities due to their abundance. The comparison of the bioactivity space showed that only eight pairs of plants with similar bioactivity spaces were phylogenetically related at either genus, family or order level. The observations indicate that the common MOAs for a large number of TM plants are not related to their phylogenies. The annotation of the bioactivity space with target classes revealed kinase was the most significant target class in all TMs. Four, three and six targets classes were observed to be more significant in TCM, Ayurveda and Malay TM respectively suggesting TMs could exhibit differential MOAs. By comparing chemical and bioactivity space, these approaches provide a closer understanding of the common and differential MOAs of different anti-cancer plants used by TCM, Ayurveda and Malay TM for treating cancers.

Chapter 4:

Evaluating Synergistic Pairwise Combinations of Shexiang Baixin Pill (SBP) for Coronary Heart Disease from Network Topology

4.1 Introduction

The World Health Organization (WHO) has projected that cardiovascular disease (CVD) will be the leading cause of death globally by 2030 with 23.6 millions deaths (250). A staggering statistic in 2012 already showed that at least 42% of the deaths relating to cardiovascular disease were contributed by coronary heart disease (CHD). This is also projected to increase (250). CHD is a disease pertaining to the blood vessels supplying the heart muscle, whose clinical symptoms include angina, myocardial infarction, sudden death and chronic heart failure (251). The pathophysiology of CHD has been generally described as an imbalance of supply and demand of oxygen in the heart muscle, which is also known as myocardial hypoxia, which eventually leads to a sudden reduction in coronary flow (252). To reduce the risk and fatal consequences of CHD, several drugs have been prescribed to patients, such as anti-platelet drugs, angiotensin converting enzyme inhibitors, beta-blockers, statins and calcium-channel blockers, in addition to a surgery option, coronary angioplasty (253). Despite the significant progress that has been observed in the treatment of CHD, there are a subset of patients who are not suitable for conventional treatments either due to acquiring complex diseases or having side effects to the current drugs (253, 254). Hence, there is an urgent need to develop new therapeutic tools to overcome the limitations of current treatments.

Neovascularization through angiogenesis provides a natural repair mechanism for CHD. The mechanism aims to improve blood flow by promoting the formation of new blood vessels to resupply blood flow to the hypoxic cardiac tissue (255). In addition to the development of therapeutic angiogenesis (255), traditional medicine, such as TCM, also offers a natural remedy for treating CHD. Shexiang Baixin Pill (SBP) is one of the widely used TCM formulations for CHD, which has been claimed to promote vasodilation and angiogenesis in the heart (82). The therapeutic effects of SBP were recently reported by a meta-analysis of 26 studies of 2,634 cases collected between 2010 and 2016, which showed SBP improved electrocardiogram rate of patients with fewer side effects than isosorbide dinitrate for angina pectoris (256). The therapeutic effects of SBP are derived from seven different Chinese *materia medica*, namely *Moschus* (musk deer), *Panax Ginseng* (ginseng plant), *Calculus Bovis* (dried gallstone of cattle), *Cortex Cinnamomi* (cinnamon twig), *Styrax* (resin of rose maloes), *Venenum Bufonis* (toad venom) and *Borneolum Syntheticum* (borneol) (257). The determination of the SBP compounds has been reported in a few studies (257-259).

The studies have identified at least 17 compounds in the formulation, which are regarded as the SBP bioactive compounds. Consequently, the identification of SBP bioactive compounds allow mechanistic studies to understand the therapeutic effect of SBP.

The therapeutic effects have been investigated in a few *in vivo* studies. Xiang et al. reported a comparative serum metabolomics study of experimentally induced myocardial infarction (MI) in rats that were treated with SBP, a multicomponent formulation called SAHRA and five single agents (260). The study revealed that the SBP-treated group showed better therapeutic effects than the remaining two groups by regulating three out of four pathological processes that are involved in MI development (260). In a follow-up study, experimentally induced MI in rats was treated with seven SBP compounds individually and an all in one combination (261). The results showed that the physiological parameters of the combination were closer to the control group than the monotherapy (261). The interactions between the SBP compounds might contribute to the therapeutic effects (261). Jiang et al., demonstrated that borneol, one of the active SBP components, enhanced the plasma concentration of four types of ginsenoside after oral administration to male Sprague-Dawley rats (262). Given the collection of reports, the studies support the claim of the therapeutic effects of SBP, which were achieved by the interactions between the compounds.

However, the MOAs underlying these therapeutics effects remained unexplained. The elucidation process is challenging if the compounds are to be tested using the present evaluation paradigm for single chemical compounds (263). For this reason, a combination of two compounds was selected to understand the MOAs. The study of compound combination has become a more common approach in drug discovery from the understanding of a complex disease that involves multiple targets and pathways (264). The effects of compound combinations can either be synergistic, additive, antagonistic or none, where synergy is the desired effect. A synergistic effect results from two compounds that produce an effect greater than the sum of their individual effects. The decision process to determine synergy of a combination could be measured using three popular strategies, which are Highest single agent (HSA) (265), Loewe additivity (266) and the Bliss independence model (267). The HSA defines the expected synergy effect (Y_{HSE}) is higher than the maximum effect of individual compounds (y_1 and y_2), $Y_{HSE} = \max(y_1, y_2)$. Therefore, synergy is determined if an experimental combination shows effect higher than y_1 or y_2 . The Loewe additivity quantifies synergy if the Combination Index (CI) < 1. The CI value can be measured from the equation below.

$$CI = \frac{a}{A} + \frac{b}{B}$$

Equation (4.1)

(a , b) are the doses of compounds 1 and 2 in a combination and (A , B) are the single doses of compounds 1 and 2. The Bliss independence model states that two compounds exert their effects independently, which the expected synergy effect (Y_{BLISS}) of two compounds (Y_1 and Y_2) can be measured from $Y_{\text{BLISS}} = Y_1 + Y_2 - Y_1Y_2$. Hence, synergy is determined if an experimental combination shows an effect higher than Y_{BLISS} .

Although, the identification of synergistic compound combinations is achievable in high-throughput screening, usually only a small number of combinations are identified (264, 268, 269). Therefore, computational approaches have been preferred to quantify, filter and rank compound combinations rapidly for guiding experimental investigations (264, 268, 269). One such computational approach is predicting compound combinations based on the PPI interaction network (269). For example, Li et al., generated a disease-specific network using PPI from HPRD and KEGG (270). Based on the network, an algorithm, Network-target-based Identification of Multicomponent Synergy (NIMS) was developed to evaluate synergistic combinations calculated from the topological relationships of the combinations of targets on a disease-specific network and their phenotypes similarities. The combinations with high synergy score were validated experimentally and their MOAs were characterised from the constructed network. Synergy evaluation was also investigated using drugs. Huang et al. established a systematic tool, DrugComboRanker (271). The tool ranked drug synergy combinations measured from the topological properties of a network that was generated using genomic profiles (271). The targets of the top-ranked synergy drugs were then mapped onto a disease-signalling network that could elucidate the drugs' MOAs that were supported by the literature. Hence, it appears that the topological properties of a PPI network maybe able to prioritise synergistic combinations and eventually elucidate the MOAs.

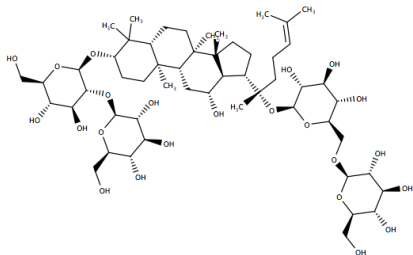
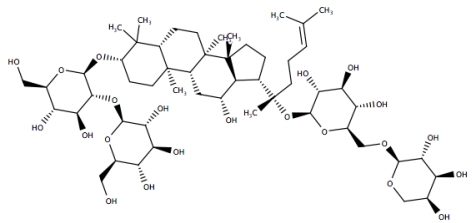
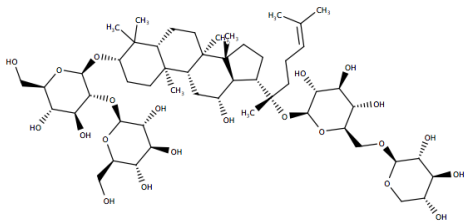
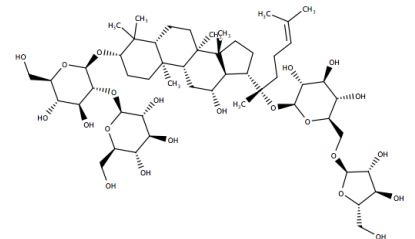
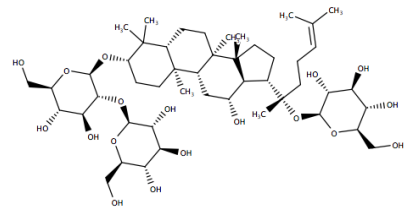
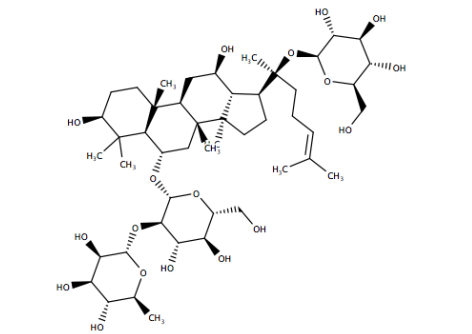
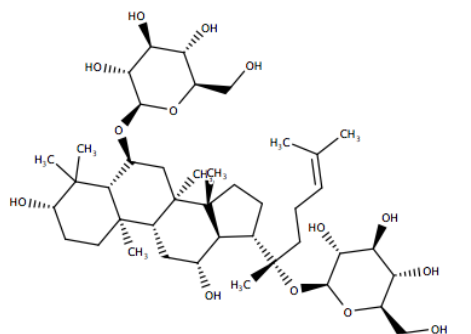
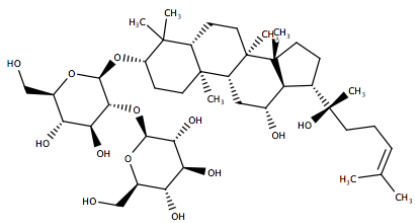
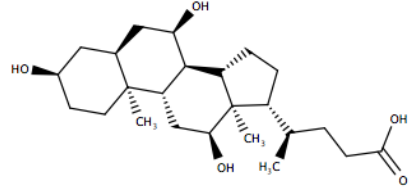
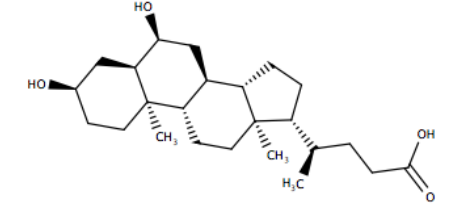
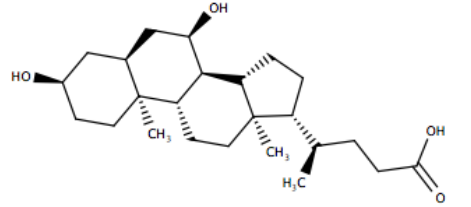
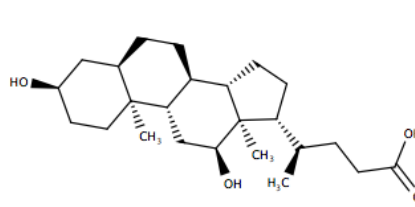
To address the question of synergistic combinations of SBP compounds, this study aimed to identify the synergistic compound combinations of pairwise combinations of 22 SBP compounds by calculating the synergy scores from a PPI network. The 20 top-ranked and bottom-ranked combinations were validated using Human Umbilical Vascular Endothelial Cells (HUVEC). The Bliss synergy scores were measured to determine the robustness of the synergy score calculation. In the end, the MOAs of the top-ranked combinations were described based on the connections of the predicted targets in the network. Hence, this method is able to suggest the MOAs of two SBP compounds to explain the claimed synergistic effects.

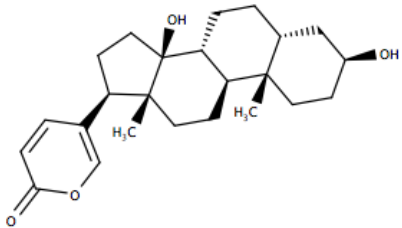
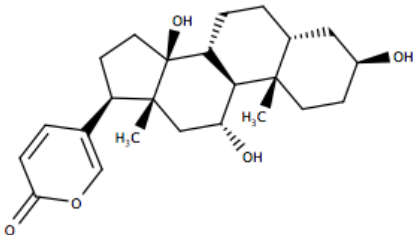
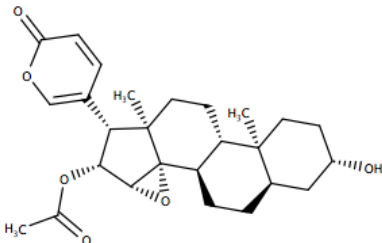
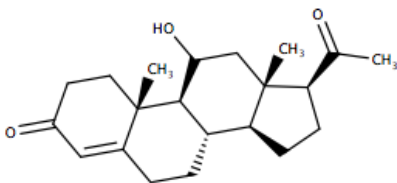
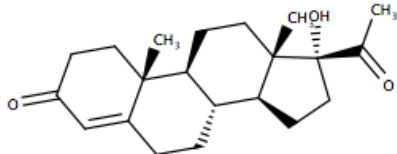
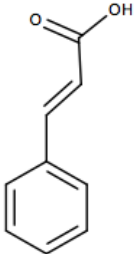
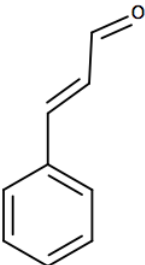
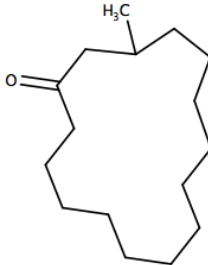
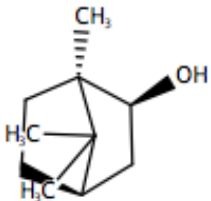
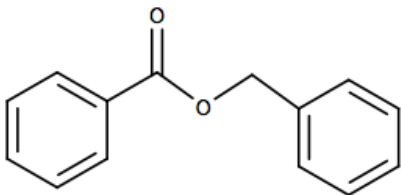
4.2 Materials and methods

4.2.1 Dataset preparation

The 22 SBP compounds (**Table 4.1**) were suggested from an on-going study of SBP by Dr Tai-Ping Fan (Department of Pharmacology, University of Cambridge). The structures of the compounds were downloaded from PubChem (110). The compounds in SMILES format were uploaded to Standardizer (210) to standardise the compounds prior to *in silico* target prediction. Six pre-selected actions were specified in the configuration file; ‘Remove Fragment’, ‘Remove Explicit Hydrogens’, ‘Neutralize’, ‘Clean 2D’, ‘Mesomerize’ and ‘Tautomerize’. The standardised SBP compounds were saved in SMILES format.

Table 4.1: The structures of 22 SBP compounds, which similar structures are arranged next to each other. PCID stands for PubChem ID.

<p>Ginsenoside Rb1 (PCID 9898279)</p> 	<p>Ginsenoside Rb2 (PCID 6917976)</p> 	<p>Ginsenoside Rb3 (PCID 12912363)</p> 	<p>Ginsenoside Rc (PCID 12855889)</p> 
<p>Ginsenoside Rd (PCID 11679800)</p> 	<p>Ginsenoside Re (PCID 441921)</p> 	<p>Ginsenoside Rg1 (PCID 441923)</p> 	<p>Ginsenoside Rg3 (PCID 9918693)</p> 
<p>Cholic acid (PCID 221493)</p> 	<p>Hyodeoxycholic acid (PCID 5283820)</p> 	<p>Chenodeoxycholic acid (PCID 10133)</p> 	<p>Deoxycholic acid (PCID 222528)</p> 

<p>Bufalin (PCID 9547215)</p> 	<p>Gamabufotalin (PCID 259803)</p> 	<p>Cinobufagin (PCID 11969542)</p> 	<p>11-hydroxyprogesterone (PCID 92750)</p> 
<p>17-hydroxyprogesterone (PCID 6238)</p> 	<p>Cinnamic acid (PCID 444539)</p> 	<p>Cinnamaldehyde (PCID 637511)</p> 	<p>Muscone (PCID 10947)</p> 
<p>Borneol (PCID 6552009)</p> 	<p>Benzyl benzoate (PCID 2345)</p> 		

4.2.2 Target prediction

The standardised compounds were run on an *in silico* target prediction algorithm that was modeled using the Random Forest (211). Briefly, the model contained a training set of active compounds extracted from ChEMBL v.21.0 (28) and inactive compounds extracted from the PubChem Compound and PubChem BioAssay databases (212). In total, the model contained 5,888,615 ligand-target pairs of actives and inactive associate with 1,651 targets. The active training set were active compounds with reported activities ($K_i/K_d/IC_{50}/EC_{50}$) of lower than 10 μ M with a CONFIDENCE_SCORE of 5 or greater for ‘binding’ or ‘functional’ human protein assays (28). The inactive training set were inactive compounds of assays submitted by researchers to PubChem Compound and PubChem BioAssay databases (28). The descriptor of a molecule was generated based on ECFP_4 circular Morgan-type fingerprints (48) with a 2,048 bit length, hence stereochemistry of compounds was excluded. The algorithm calculates the probability of an orphan compound belonging to a target class by considering both chemical features of active and inactive compounds against target classes, giving a more holistic perspective of chemical features, which contribute towards or against bioactivity.

To predict targets of TM compounds, the Random Forest (RF) classifier can be briefly described as follows. First, a Scikit-learn (Version 17) RF classifier of 100 trees, with the number of features and max depth set to ‘auto’ and the ‘class_weight’ set to ‘balanced’, was trained using the binary matrix of the active and inactive compound-target fingerprints on a per-target basis (213). Models generate a probability representing the number of trees in the forest that predict a target as active, represented as $p(activity)$. Platt scaling (214), a calibration method in machine learning, is performed to convert the classification from the random forest into the probabilities of the prediction being correct, which is known as True Positive Rate (TPR).

4.2.3 The TPR of predicted targets

The TPR value was selected by assessing the distribution of a fraction of targets predicted for each TPR interval of 0.1 between 0 and 1. A curve of the fraction of targets’ predicted TPR intervals was evaluated to define a TPR threshold of each SBP compound (**Figure 4.1**). In general, the fraction of targets predicted decreases as the threshold value increases. A threshold value of 0.65 was selected for all SBP compounds because at least four targets were predicted for ginsenoside compounds. The selected threshold value is a reasonable trade-off between false-negative predictions and the number of predicted targets compared to larger threshold values. By using the selected TPR threshold, the target prediction algorithm was performed, which produced a binary matrix of predicted targets.

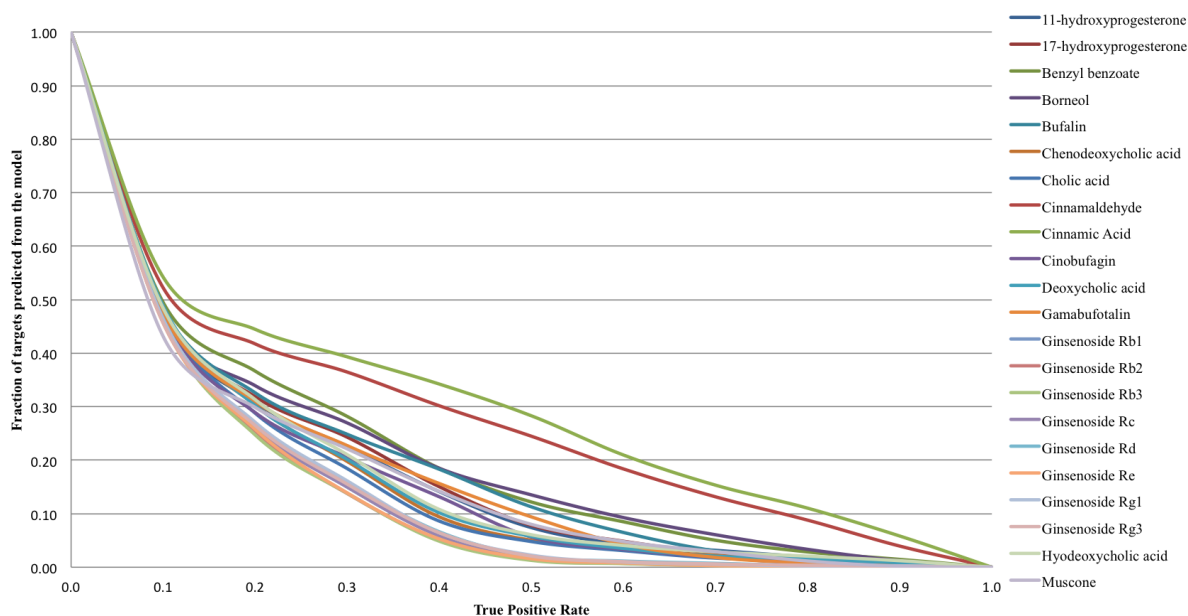


Figure 4.1: The curves of the fraction of targets predicted decreases when TPR values increases. A TPR threshold value of 0.65 was selected in this analysis because at least four targets were predicted for ginsenoside compounds.

4.2.4 Generating a representative angiogenesis and CHD network

To generate the PPI network, angiogenesis and CHD annotated genes were extracted from six databases (**Table 4.2**). By using the keyword “angiogenesis”, the annotated genes were extracted from UniProt (272) and Comparative Toxicogenomics Database (CTD) (273). The annotated genes from Gene Ontology (GO) were retrieved using an identifier, GO:0001525 (274). The CHD annotated genes were extracted from UniProt (272) and Target Validation Platform (TVP) (215) by using the keyword “coronary heart disease”. In CTD (273), the CHD annotated genes were retrieved from six CTD identifiers, D006333 Heart Failure, D001145 Arrhythmias, Cardiac, D006323 Heart Arrest, D017202 Myocardial Ischemia, D006341 Heart Rupture, and D009202 Cardiomyopathies. To prioritise the extracted genes from CTD, the top 1,000 angiogenesis annotated genes were selected based on the Inference Score. In another approach, the CHD annotated genes from six identifiers were combined and the top 1,000 CHD annotated genes were selected based on the Inference Score. The CHD annotated genes of TVP (215) were filtered based on an Association Score of at least 0.7, which at least 80% of genes were included. All annotated targets from UniProt (272) and GO (274) were kept. For the purpose of this study, all genes were regarded to encode their complementary proteins.

Table 4.2: The numbers of annotated genes of angiogenesis and CHD, which were extracted from six databases and the final numbers of genes to construct the network.

Angiogenesis		Coronary Heart Disease	
Database	Number of Genes	Database	Number of Genes
Gene Ontology (GO:0001525)	404	Target Validation Platform	727
Uniprot	494	Uniprot	935
Comparative Toxicogenomics Database	1,000	Comparative Toxicogenomics Database	1,000
Numbers of genes for network generation	407	Numbers of genes for network generation	304

Initially, two separate networks were generated for angiogenesis and CHD by selecting proteins that were consistently found in at least two databases. The PPIs between two proteins were retrieved from Reactome (275) before merging them into a new network. Singnalink2.0 database (66) was also incorporated to further expand the network. The selected layer of Signallink2.0 (66) was Pathway Member composed of 1,086 nodes and 1,226 edges. Cytoscape v3.3.0 (276) was used in each step of the network generation. In a network, the protein is known as a node while the link between two nodes is known as an edge. Self-loops and multiple edges between the same pairs of nodes were removed. Only the first-degree neighbours of SBP targets were selected for generating the final network.

4.2.5 Network properties analysis

The properties of the network were assessed based on the topology such as degree distribution, average clustering coefficient and average shortest path length. Degree (k) is the number of edges a node has. Degree distribution of a network was calculated from the frequency (f) of each degree from the total number of nodes (n):

$$\langle k \rangle = \frac{f_k}{n}$$

(Equation 4.2)

The degree distribution was analysed based on the annotation of the node attributes in **Table 4.3**. The boxplot each node's attribute was created using boxplot() function of stats package in R (114).

Real biological networks exhibit a scale-free behaviour. To determine this property, the average clustering coefficient and average path length of the network were compared to 1,000 random networks of the same number of nodes and edges and the edges were permuted. The random networks were generated by randomising the degree of the original network using degree.sequence.game() function of igraph (277) package in R (114). The clustering

coefficient is defined as sets of three nodes which are connected to each other. This measure provides information on how neighbours are connected in the network, where the value is between 0 and 1. The average clustering coefficient $\langle C \rangle$ of a network can be quantified by:

$$\langle C \rangle = \frac{1}{n} \sum_{i=1}^n C_i$$

Equation (4.3)

Here, C_i is the ratio of number of triangles connected to node i to the number of triples (both open and closed triangles) centred on node i . The sum of C_i is averaged by the total number of nodes (n).

The average path length is defined as the number of average shortest steps for all possible pairs of nodes in the network, in which a shorter average path length is more desirable. The average path length $\langle l \rangle$ of a network can be quantified by:

$$\langle l \rangle = \frac{1}{n(n-1)} \sum_{i \neq j} d(n_i, n_j)$$

Equation (4.4)

Here, $d(n_i, n_j)$ is a measure of the shortest distance between node i and node j . The sum of the distance for all possible pairs of nodes is averaged by the number of all possible pairs of nodes.

4.2.6 Biological process analysis

The functional pathway enrichment of the SBP predicted targets and all proteins in the network was measured by ClueGO (278) by selecting GO Biological Process as the background set. The statistical test used for the enrichment was based on a two-sided hypergeometric option with a Bonferroni correction, a p-value less than 0.001. A fusion term was used with a defined kappa score of 0.65. The level of specificity of GO terms was selected between 3 and 13.

4.2.7 Formulating the synergy score of compound combinations

In total, there were 231 unique pairwise combinations generated from 22 SBP compounds. To measure a synergy value of a combination, three elements were integrated. The three elements were topological properties, pathway dissimilarity and average node distance. The synergy score was calculated by using *Equation (4.5)*. The equation was modified from a previous study by Li et al., (270) by adding two new elements, pathway dissimilarity and mean distance. The Synergy Score (SS) of compound i and compound j was a product of a

Topology Score (TS) and a Pathway Dissimilarity Score (PDS) divided by a Mean Distance Score (MDS).

The TS was derived from two topological features and a defined node attribute of the predicted targets for compound i and compound j as described below.

$$TS(i, j) = \frac{1}{2} \left[\frac{\sum_i (Deg \times Betw \times WA)^{-1/3}}{n_i} + \frac{\sum_j (Deg \times Betw \times WA)^{-1/3}}{n_j} \right]$$

Equation (4.2)

From the above equation, TS was a harmonic average product of the TS of compound i and the TS of compound j . The TS of a compound was defined as a sum of three elements, degree (Deg), betweenness (Betw) and weight attribute (WA). The Deg and Betw were calculated based on the connections of a node in the network, while the WA was defined as whether a node was annotated as an angiogenesis protein, and/or a CHD protein. Both angiogenesis and CHD proteins were proteins that were extracted from six databases to construct the network. The scheme of the WA was described in **Table 4.3**.

Table 4.3: The weight attribute of the nodes in the network.

Node Attribute	Weight
Angiogenesis (An)	2
Coronary Heart Disease (CHD)	2
An and CHD	3
Not An and CHD	1

The sum of product of Betw, Deg and WA was normalised using taking the reciprocal of its cube root and the final TS was averaged by the number of predicted SBP targets, n , for that compound.

The second element of the SS score was to calculate the pathway dissimilarity between two compounds. The pathways were derived from the gene functional enrichment calculation in the network analysis. A matrix of frequency of nodes to a pathway was created. Pearson coefficient correlation (PCC) was used to calculate the pathways similarity. The calculated value was subtracted by 1 to get the PDS.

$$PDS(i, j) = 1 - PCC(i, j)$$

Equation (4.3)

Prior to calculating the PDS, a heatmap of the bioactivity space similarity of SBP compounds was assessed. The frequencies of compounds of per predicted target were normalised. The dissimilarity of distances between the normalised frequencies of SBP compounds were calculated using ‘Euclidean’ method of the “dist” function. Clustering was performed using the Ward method of the “hclust” function, where two SBP compounds were merged if the sum of the squared Euclidean distance was minimal. The heatmaps were generated using the “heatmap” function. All the aforementioned functions were available in the stats package of R (114).

The third element, MDS, was calculated from the sum of the shortest distance between all pairwise pairs of targets of compound i and compound j (iCj), which was averaged by the number of pairwise targets, n_{iCj} .

$$MDS(i, j) = \frac{\sum_{iCj} d_{min}}{n_{iCj}}$$

Equation (4.4)

The three elements were combined into the SS equation as below:

$$SS(i, j) = \frac{TS(i, j) \times PDS(i, j)}{MDS(i, j)}$$

Equation (4.5)

A Spearman Rank Correlation Coefficient (SRCC) was used to assess the relationship between the scores of the three elements and the synergy scores. SRCC was measured using the “cor” function of R the stats package (114).

4.2.8 Robustness of Synergy Score evaluation

To evaluate the robustness of the *Equation (4.5)*, the SS of a combination of the original network were compared to SS of a background distribution by computing the SS from 1000 random graphs of the same number of nodes and edges. The nodes of each compound were randomly selected based on the number of targets predicted. The enrichment of a SS from the original network was measured based on Estimation Score (ES).

$$ES = \frac{C}{1000}$$

Equation (4.8)

Here, C is defined as where the frequency of SS in the 1000 random graphs is smaller than SS in the original network. The absolute frequency obtained is then divided by total number of

random graphs giving a value between 0 and 1. A combination is considered enriched when $ES \leq 0.01$.

4.2.9 Synergistic evaluation of combinations in HUVEC

In order to validate the predicted SS, five combinations from the 20 top-ranked and five combinations from the bottom ranked were treated in HUVEC. HUVECs were cultured in Promocell's Endothelial Cell Growth Medium 2 containing 2% FBS, human Epidermal Growth Factor, human Basic Fibroblast Growth Factor, Insulin-like Growth Factor, human Vascular Endothelial Growth Factor 165, Ascorbic Acid and Heparin (EGM-2). The HUVECs were incubated at 37°C in humidified air containing 5% carbon dioxide. HUVECs were seeded in 96-well plates at a density of 2,500 cells/well in 100µL of EGM-2. 24 hours later, the media was aspirated and appropriate compound concentrations diluted in Endothelial Cell Basal Medium 2 (EBM-2, Promocell) were added to the relevant wells (total volume in each well = 100µL). After 72hr of incubation, 10µL of Cell Counting Kit-8 (CCK-8, Dojindo) was added to each well. CCK-8 utilises Dojindo's highly water-soluble tetrazolium salt (WST-8). Dehydrogenases within cells reduce WST-8 to produce orange formazan. The amount of formazan dye generated is proportional to the number of living cells. The plates were incubated for 1-4 hours at 37°C. A MultiSkan Ascent Plate Reader measured absorbance at 450nm. The experimental work was performed by Dr Ranjoo Choi.

4.2.10 Bliss independence model measurement

The combination response (4x4) matrix contained percentages of cell growth and at zero concentration, the cell growth was equal to 100%. Each value in the matrix was subtracted from 100 to produce percentage cell growth. The combination matrix was plotted as a heatmap. A single dose response curve was constructed prior to calculating the Bliss independence model. The single dose response curve was fitted using 4-parameter log-logistic function (4-PL). The Bliss score of each combination was calculated by averaging (3x3 matrix) the difference between the actual percentage of cell growth and the expected percentage of cell growth. The combination interaction landscape was visualised in a 3D contour plot. All the calculations and figures of this section were carried out using the synergy finder package (279) in R (114).

4.3 Results and Discussions

4.3.1 Analysis of pathways associated with SBP predicted targets

Prior to the network construction, the analysis of the SBP predicted targets shows that the predicted targets are associated with GO biological processes (BP) that are related to either angiogenesis or CHD (**Table A.4.1**). Such biological processes are mitogen activated protein kinase (MAPK)-related pathways (GO:0000187, GO:0032872, GO:0043405), angiogenesis (GO:0001525), ERK1 and ERK2 cascade (GO:0070371), T-cell activation (GO:0042110), lymphocyte differentiation (GO:0030098), blood vessel morphogenesis (GO:0048514), and epidermal growth factor receptor signalling pathway (GO:0030098). Numerous studies have demonstrated that angiogenesis and/or CHD affected these biological processes.

For example, atherosclerotic process cause thrombosis in arteries that leads to MI in CHD (280). The over accumulation of cholesterol on the artery walls induces cholesterol crystal formation that can activate an associated inflammasome such as cytokines (280). Inflammatory cytokine production is one of the key features in CHD. T cell activation and lymphocyte differentiation prevent the accumulation of these pro-inflammatory cytokines, hence suppressing the atherosclerotic state (281). The cross talk between inflammation and angiogenesis has also been reported, where the infiltration of T cells into autoimmune lesions induces pathogenic angiogenesis by secreting VEGF (282). MAPK-related pathways have been widely studied for their associations with cardiovascular diseases. Four subfamilies, ERK1/2, JNK, p38, and ERK5 of MAPK, have been characterised for their involvements in cardiac development, function and diseases (283). One of the subfamilies, ERK1 and ERK2 cascade, is among the enriched GO BP terms. The up-regulation of ERK 1/2 MAPK phosphorylation induces anti-inflammatory cytokine, interleukin-10 (IL-10), to prevent tumour necrosis factor-alpha (TNF-alpha) from increasing oxidative stress and apoptosis (284). The enrichment of GO terms, angiogenesis (GO:0001525), blood vessel morphogenesis (GO:0048514), and epidermal growth factor receptor signalling pathway (GO:0030098) are indicative of significant angiogenesis-related proteins being part of the network. Altogether, these biological processes indicate that SBP compounds were predicted to modulate targets of the pathways that are important in either angiogenesis or CHD.

4.3.2 Evaluating topological properties of the network

The final representative network of angiogenesis and CHD (**Figure 4.2**) was composed of 2,371 nodes and 16,336 edges, of which 396 nodes were angiogenesis and CHD associated genes. The 2,371 nodes were annotated to 332 enriched GO biological processes (**Table A4.2**). A total of 259 SBP predicted targets were mapped onto the final network, in addition

to 13 targets from the literature that were associated with the MOA of SBP compounds. For clarity, the 13 targets (**Table A4.3**) were also regarded as SBP predicted targets for the rest of the analysis.

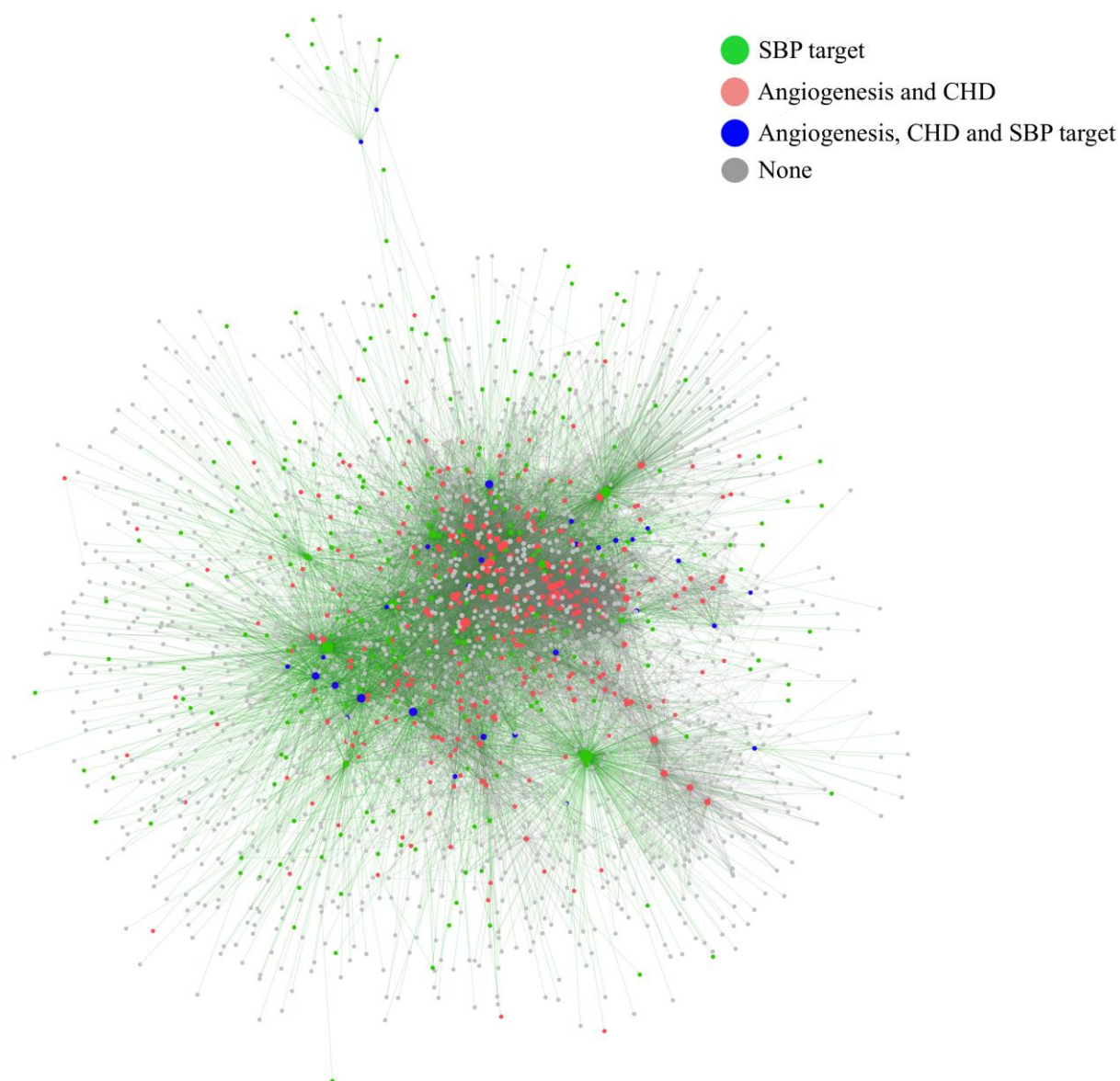


Figure 4.2: The visualisation of the representative network of CHD and angiogenesis, which the angiogenesis and CHD annotated nodes (blue and red circle) are mostly aggregated in the middle of the network. The observation suggests that these two conditions are in the biological context of the network. The SBP targets and their degree (green circle and line) are observed to link many of the nodes suggesting the targets could provoke the desired therapeutic effects.

The topological properties of the network were first evaluated by its degree distribution (**Figure 4.3**) (285). The degree distribution based on the nodes' attributes shows that the median values of two groups, angiogenesis and CHD annotated nodes and angiogenesis, CHD and SBP annotated nodes, are higher than the remaining two groups and the median values are closer to the upper quartile. The observation indicates that many of the annotated angiogenesis/CHD proteins are highly linked nodes suggesting that the two conditions are governing the conformation of the network. The highly linked nodes are also known as hub proteins. The associations of the top ten hub proteins to either angiogenesis or CHD were described in **Table 4.4**. The results demonstrate that six hub proteins were predicted to be modulated by the SBP compounds. Despite the low median value of the SBP predicted targets, some of the SBP targets are observed to be distributed outside the upper quartile. The observation demonstrates that the SBP compounds were predicted to modulate some of the hub proteins, which could be responsible for producing the desirable phenotypes. In addition to degree distribution, the properties of the network were analysed by measuring its average clustering coefficient, average betweenness and average path length. These properties were compared to 1,000 random graphs with the same number of nodes and edges as can be seen in **Table 4.5**. The higher average clustering coefficient and shorter average path length of the original network than the random graphs demonstrated that the network is a scale-free network.

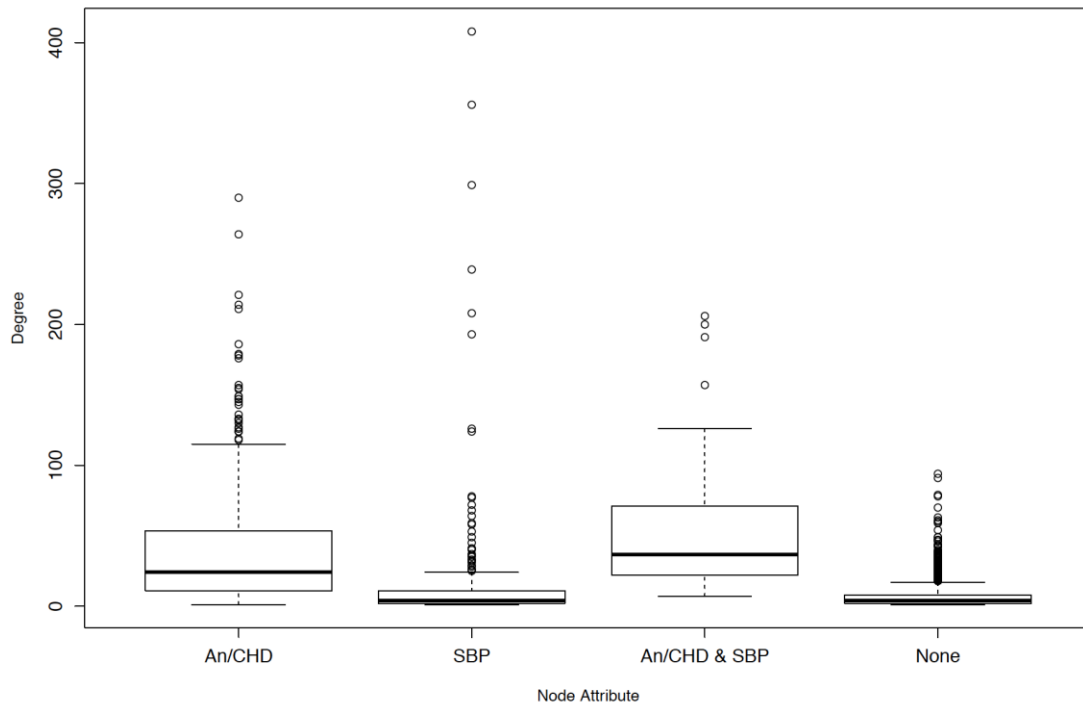


Figure 4.3: The degree distribution of different node attributes in the network. Many of the SBP predicted targets and angiogenesis and CHD annotated proteins are highly linked nodes. The observation indicates that the two conditions are governing the conformation of the network.

Table 4.5: The comparison of topological properties between a representative network of angiogenesis and CHD and 1,000 random graphs.

Properties	Original network	1,000 random networks
Number of nodes	2,371	2,371
Number of edges	16,336	16,336
Average clustering coefficient	0.136	0.00581 ± 0.00029
Average path length	3.072	3.246 ± 0.00087

Table 4.4: The top ten hub proteins in the representative network of angiogenesis and CHD. Six proteins were predicted to be modulated by SBP compounds.

Symbol	Protein Name	Attribute	Degree	Protein Family	Supporting literature
CTNNB1	Catenin beta-1	SBP/ Angiogenesis	408	Beta-catenin	It was reported that the overexpression of CTNNB1 enhances endothelial cells proliferation through cell cycle propagation and induces vascular endothelial growth factor (VEGF) in skeletal myocytes (286)
JUN	Transcription factor AP-1	SBP/ Angiogenesis	356	Jun	The activation of JUN has been showed to mediate VEGF-induced endothelial cells proliferation and migration (287).
RHOA	Transforming protein RhoA	SBP/ Angiogenesis	299	Rho	RHOA has been documented to activate VEGF-induced endothelial cells (288).
MAPK14	Mitogen-activated protein kinase 14	Angiogenesis	290	Kinase	MAPK14 is also referred as p38 MAPK, which the inhibition of p38 MPK has been reported to enhance angiogenesis in VEGF-induced cell proliferation (289).
PIK3CA	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform	Angiogenesis	264	Kinase	The PIK3CA has been frequently reported in cancer pathogenesis, which the mutation of PIK3CA induces angiogenesis and malignant cell growth (290).
RELA	Transcription factor p65	SBP/ Angiogenesis	239	Not available	The activation of RELA has been demonstrated to either promote or impair angiogenesis through matrix metalloproteinase (291).
AKT1	RAC-alpha serine/threonine-protein kinase	Angiogenesis	221	Kinase	AKT1 has been documented to exhibit pro-angiogenic and anti-angiogenic effects under different conditions (292).
SHC1	SHC-transforming protein 1	Angiogenesis	214	Not available	It has been reported that the activation of SHC1 induces the downstream pathways of epidermal growth factor receptor (EGFR) (293).
EGFR	Epidermal growth factor receptor	SBP/CHD	211	Kinase	In addition to EGFR's implication in angiogenesis, the protein has been reported to regulate blood pressure by either promoting vasodilation or vasoconstriction (294).
PTK2B	Protein-tyrosine kinase 2-beta	SBP/ Angiogenesis	208	Kinase	The activation of PTK2B has been implicated in wound healing by stimulating the Src family kinase/EGFR signalling pathway (295).

4.3.3 Comparison to previously published network

A few studies of biological network generation of angiogenesis or CVD have been reported in the literature. These studies used different methods in order to generate the biological network. In this section, we compare the 2,371 genes of the constructed network to the genes of the three angiogenesis networks and one CVD network that were previously published (Table 4.5). The comparison discovers that in many cases, the annotated genes that were used in the current network were also part of the networks published previously.

Table 4.6: The comparison of overlapped genes of angiogenesis/CHD related networks discovered from previous studies.

Number of nodes	Number of edges	Method	Number of overlapped genes
1,233	5,736	Extracted angiogenesis genes from multiple sources; Gene Ontology, SABiosciences, and Gene Cards and used GeneHit to expand the network of angiogenesis dynamically (296).	382 (580 angiogenesis genes)
362	1,195	Extracted angiogenesis genes from multiple sources to generate an angiogenesis network ; 1. Database (Gene Ontology, Reactome, Uniprot), 2. computational and experimental tools (Protein lounge, Ingenuity Pathway, Agilent Literature search), 3.commercial experimental sets (G-Arrays) (297).	57 (79 angiogenesis genes)
unknown	262	Used ClusteEx method: identifies groups of differentially expressed genes and connects the genes' products using shortest path in VEGF stimulated human umbilical vein endothelial cells (HUVECs) to represent inflammation and angiogenesis models (298)	51 (92 angiogenesis genes)
3,412	13,164	Used gene co-expression data to extract related CVD genes and Gene Ontology to develop the interactions between genes (299)	269 (705 CVD genes)

4.3.4 The modelling of synergy score measurement and the score distribution

The hypothesis behind the synergy score measurement was to relate the topological relations between two sets of predicted targets of a two-compound combination. These predicted targets were assumed to be part of different pathways with a small distance between them. The TS measurement considered the importance of predicted targets based on their connections and attributes in the network. The negative exponential function was utilised to normalise the different scales of betweenness, degree and weigh attribute. The sum of the products of the three scales was divided by the number of predicted targets due to the high

variance from only six to 180 predicted targets. Therefore, TS gave the results that stronger desired effects could be achieved from a set of predicted targets with higher connections and annotated with high-value attributes.

The PDS was derived from the analysis of the similarity of predicted SBP targets. Many of the SBP compounds were shown to be structurally similar (**Table 4.1**), for example, six types of ginsenoside and four cholic acid derivatives. Based on the “Molecular Similarity Principle” (41), these compounds are likely to modulate a similar set of targets. The heatmap of **Figure 4.4** shows that structurally similar compounds were predicted to modulate similar targets, and these compounds are clustered together. Only a few compounds are structurally different such as muscone, benzyl benzoate and borneol. The similarity of the predicted targets was considered to affect similar pathways; hence compounds with high structure similarity would dominate the top ranked combinations. These observations were more likely to identify combinations with additive effects than synergistic effects because the compounds were expected to modulate the same targets. By considering dissimilar pathways, the synergy score calculations accounted for the different targets of two compounds, hence, identifying synergistic combinations. These different pathways are still functionally related as only 332 GO terms (**Table A4.3**) were enriched from 2,371 genes.

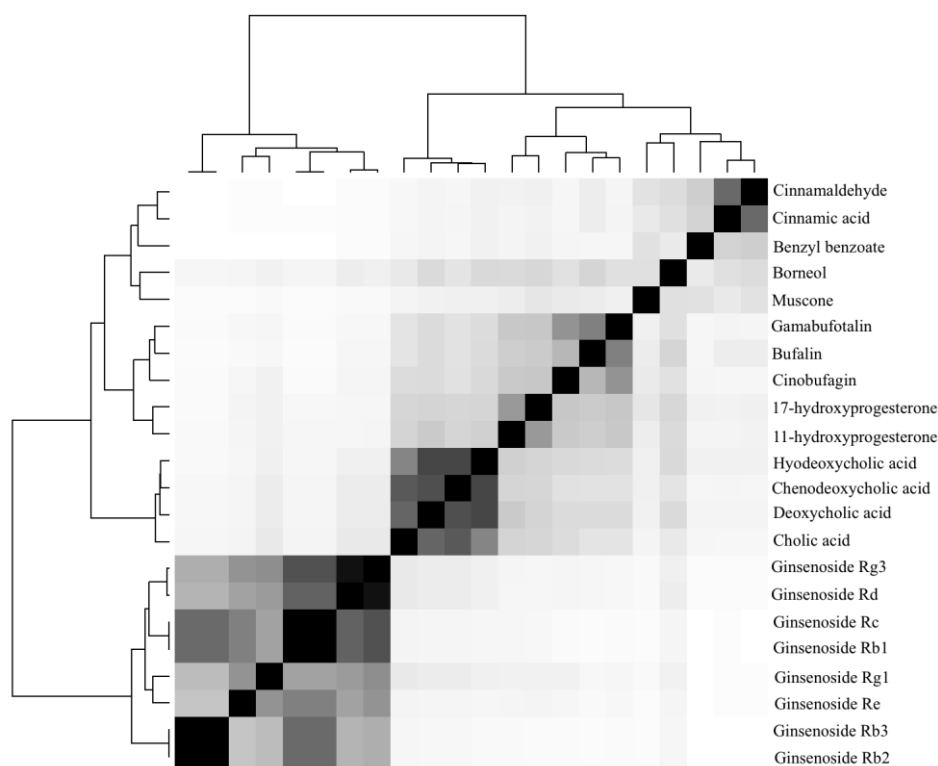


Figure 4.4: The heatmap of predicted target similarities between SBP compounds. Many of the SBP compounds are structurally similar such as ginsenoside and cholic acid derivatives. These compounds were predicted to modulate similar targets, thus, acting on the same pathways. Due to high-predicted targets similarities, the SS measurements considered dissimilar pathways annotated to the compounds.

The final element in the SS measurement was the mean distance measured from the pairwise distance of all predicted targets of two compounds. A study by Wang et al., indicates that combinatorial drugs tend to affect a smaller radius in the gene network in order to produce their therapeutic effects (300). In the study, it was found that 62% of the combinatorial drugs had a shorter effect radius of 3 and less. Given the result of the previous study, the final SS was determined by the mean distance of the predicted targets of both compounds. A combination with a smaller mean distance was estimated to show higher synergistic effects.

Spearman Rank Correlation Coefficient (SRCC) was calculated to determine the relationships between TS, PDS and MDS and SS, which is visualised in the multi-plots in **Figure 4.5**. Both TS ($r = 0.83$, $p\text{-value} < 0.001$) and PDS ($r = 0.71$, $p\text{-value} < 0.001$) show strong correlation with that of SS, whereas MDS ($r = 0.12$, $p = 0.072$) showed weak correlation with that of SS. The results suggest that the weak correlation of MDS with that of SS has prioritised combinations with targets that were close to each other (< 2.8). Therefore, the evaluation of synergy prioritises combinations that contain targets are close to each other, have many high degree nodes and belong to different pathways.

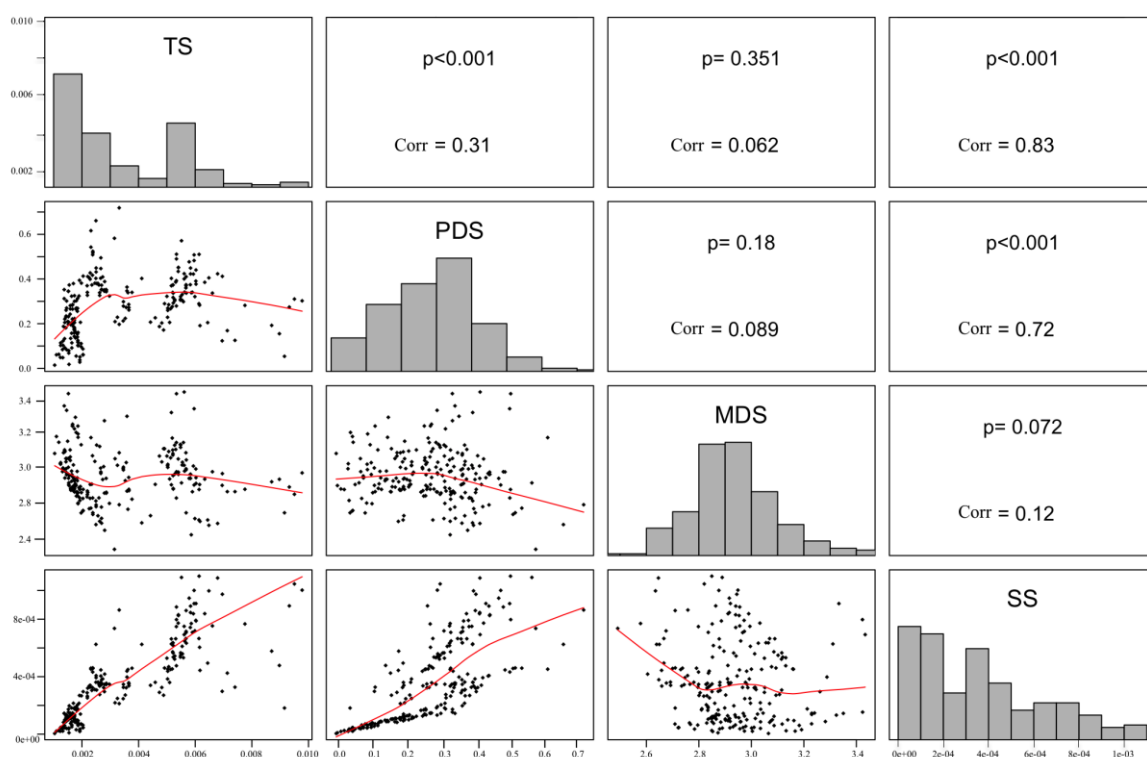


Figure 4.5: The SRCC between TS, PDS and MDS and SS. Both TS and PDS are highly correlated with that of SS and MDS is weakly correlated with that of SS. Combinations with higher SS are estimated to contain many highly linked predicted targets and both sets of the predicted targets are located in smaller radius, which belong to different pathways.

4.3.5 Robustness of Synergy Score Evaluation

It was found that all the combinations were enriched with a p-value of zero. All SS of 231 combinations measured from the original network were higher than the 1,000 SS of 231 combinations measured from 1,000 random networks. The boxplot in **Figure 4.6** compares four types of score distributions, where the scores of the random group are the average scores measured from 1,000 random networks. It is observed that all three elements influence the SS. The large mean difference of TS between the original and random groups suggesting the topological properties, degree, betweenness, and weight attribute, are important factors in evaluating synergy of a combination. Combinations with high SS measured from the original network are estimated to modulate hub proteins, which in the same combinations in the random networks, the edges of these hub proteins were permuted. The PDS distribution of the random networks is higher than the original network because the randomised targets of each combination are part of different pathways. Similar to the PDS, the MDS distribution of the random group is higher than the MDS distribution of the original group but with smaller variance. The result implies that a mean distance between set of targets of two compounds is an important parameter to emphasise synergistic effects. Hence, the synergy score evaluation is robust by emphasising combinations that are modulating high-degree nodes.

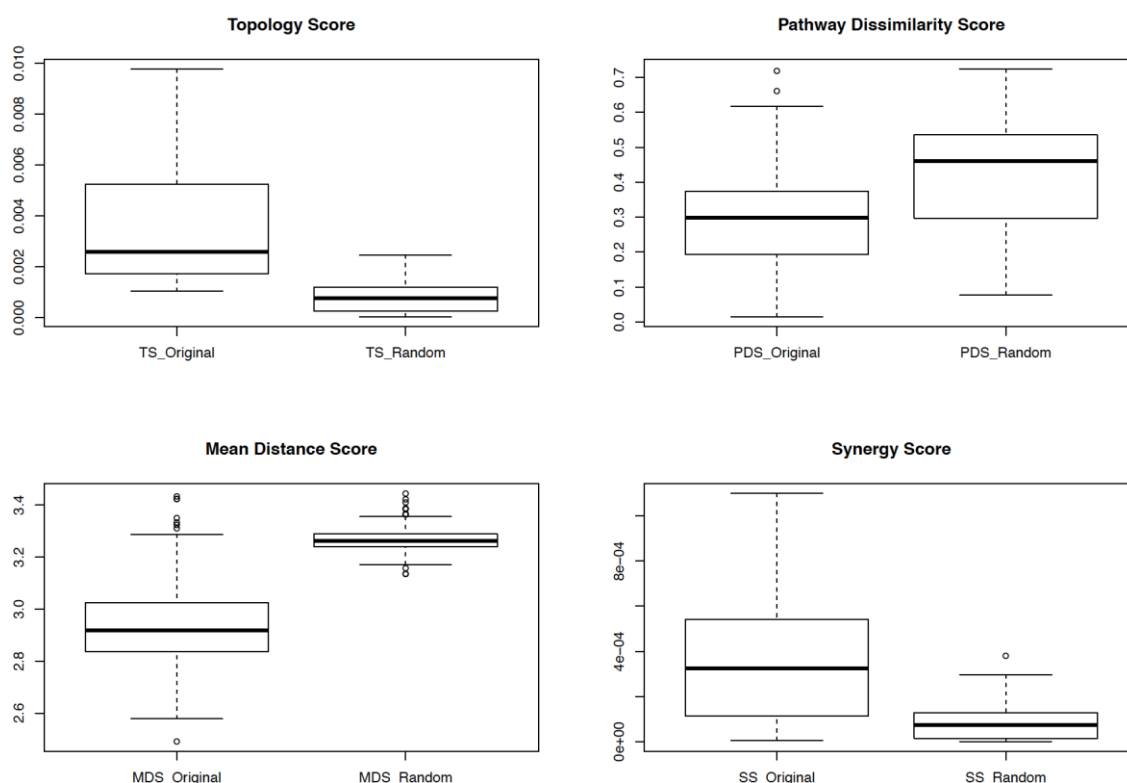


Figure 4.6: The boxplot compares four types of score distribution. The lower SS measured from 1,000 random networks are observed to be influenced by all three elements. The finding indicates that topological properties of predicted targets of a combination are significant factors for estimating synergy.

4.3.6 Predicted synergy combinations

231 combinations were arranged in descending order based on the SS (**Table A4.4**). What can be seen throughout in **Table A4.4** is that the pairs of compounds of the same or different scaffolds are randomly distributed. In most cases, combinations that contain ginsenoside are observed in the first half of the rank, while the second half of the rank contains combinations of simple structures such as borneol, cinnamic acid and cholic acid derivatives. These small molecules were predicted to modulate many targets, however, the targets were found to have less impact on the network topology. In many cases, the values of a nodes' attributes of the targets are low, resulting in a small SS of these combinations. Overall, the rank aimed to prioritise combinations that were predicted to produce synergistic effects *in vitro*.

4.3.7 Validation of predicted synergy combinations

To validate the rank of the combinations, the top and bottom 20 combinations of each category were validated using HUVEC cells by quantifying the promotion of cell growth. The top 20 combinations were estimated to show higher synergistic effects than the bottom 20 combinations. At the time of thesis submission, only five combinations from each category were validated experimentally and their Bliss scores were calculated (**Table 4.7** and **Table 4.8**). Three combinations of the 20 top-ranked combinations are combinations of ginsenoside, which are ginsenoside Rb3 and ginsenoside Rb2 (SS=9.97E-04), ginsenoside Rg1 and ginsenoside Rb2 (SS=8.64E-04) and ginsenoside Rd and ginsenoside Rb2 (SS=8.60E-04). The remaining two combinations are ginsenoside Rb3 and cholic acid (SS=1.10E-03) and ginsenoside Rb3 and 11-hydroxyprogesterone (SS=9.09 E-04). Three combinations of the 20 bottom-ranked combinations are combinations of cholic acid derivatives. The combinations are cholic acid and chenodeoxycholic acid (SS=4.51E-05), deoxycholic acid and chenodeoxycholic acid (SS=3.54E-05) and hyodexycholic acid and chenodeoxycholic acid (SS=2.52E-05). One combination contains compounds from *Venenum Bufonis* (toad venom), which is gamabufotalin and bufalin (SS=9.61E-06). One remaining combination is muscone and benzyl benzoate (SS=4.00E-04). Overall, Bliss synergy scores of the combinations in the top rank are higher than the combinations in the bottom rank except for ginsenoside Rg1 and ginsenoside Rb2. The synergistic combinations of dose-response matrices can be found in **FigureA4.1**. The MOAs of the synergistic combinations were proposed in the next section.

Table 4.7: The list of top 20 combinations, which are estimated to show synergistic effects.

Compound <i>i</i>	Compound <i>j</i>	TDS	PDS	MDS	SS	Estimation Score	Bliss score
Ginsenoside Rb3	Cholic acid	6.14E-03	5.10E-01	2.85	1.10E-03	0	6.51
Ginsenoside Rc	Cholic acid	5.51E-03	5.71E-01	2.89	1.09E-03	0	-
Ginsenoside Re	Ginsenoside Rb3	6.79E-03	4.23E-01	2.65	1.09E-03	0	-
Ginsenoside Rc	Ginsenoside Rb3	9.50E-03	3.10E-01	2.82	1.05E-03	0	-
Ginsenoside Rb3	Ginsenoside Rb1	9.77E-03	3.02E-01	2.95	1.00E-03	0	-
Ginsenoside Rb3	11-hydroxyprogesterone	5.96E-03	4.83E-01	2.88	1.00E-03	0	-
Ginsenoside Rb3	Ginsenoside Rb2	6.48E-03	4.05E-01	2.63	9.97E-04	0	13.99
Ginsenoside Rb1	Ginsenoside Rb3	6.96E-03	4.12E-01	2.94	9.74E-04	0	-
Ginsenoside Rb1	Cholic acid	5.79E-03	4.77E-01	2.92	9.46E-04	0	-
Ginsenoside Rb3	Deoxycholic acid	6.03E-03	4.40E-01	2.92	9.09E-04	0	-
Ginsenoside Rb3	11-hydroxyprogesterone	5.96E-03	5.08E-01	3.33	9.09E-04	0	7.67
Ginsenoside Rd	Ginsenoside Rb3	9.32E-03	2.74E-01	2.86	8.93E-04	0	-
Ginsenoside Rb3	Gamabufotalin	5.84E-03	4.49E-01	2.94	8.91E-04	0	-
Ginsenoside Rc	17-hydroxyprogesterone	5.34E-03	4.90E-01	2.96	8.85E-04	0	-
Ginsenoside Rg1	Ginsenoside Rb2	3.32E-03	7.19E-01	2.76	8.64E-04	0	-10.29
Ginsenoside Rd	Ginsenoside Rb2	5.68E-03	4.10E-01	2.71	8.60E-04	0	-3.71
Ginsenoside Rb3	Bufalin	5.83E-03	4.23E-01	2.92	8.46E-04	0	-
Ginsenoside Rg1	Ginsenoside Rc	6.34E-03	3.85E-01	2.91	8.39E-04	0	-
Ginsenoside Rb3	Cinobufagin	5.93E-03	4.22E-01	2.99	8.38E-04	0	-
Ginsenoside Rc	Chenodeoxycholic acid	5.41E-03	4.52E-01	2.94	8.31E-04	0	-

Table 4.8: The list of bottom 20 combinations, which are estimated to show lower synergistic effects.

Compound <i>i</i>	Compound <i>j</i>	TDS	PDS	MDS	SS	Estimation Score	Bliss score
Gamabufotalin	Cinnamic acid	1.25E-03	1.14E-01	3.01	4.73E-05	0	-
Cinnamaldehyde	Bufalin	1.33E-03	1.03E-01	2.97	4.63E-05	0	-
Deoxycholic acid	Cholic acid	2.04E-03	6.16E-02	2.72	4.61E-05	0	-
Cholic acid	Chenodeoxycholic acid	2.04E-03	6.02E-02	2.72	4.51E-05	0	-6.65
Muscone	Cinnamaldehyde	1.40E-03	9.20E-02	2.92	4.40E-05	0	-
Bufalin	Cinnamic acid	1.25E-03	1.02E-01	3.00	4.26E-05	0	-
Borneol	Gamabufotalin	1.54E-03	7.94E-02	2.89	4.22E-05	0	-
Muscone	Benzyl benzoate	1.46E-03	8.32E-02	3.03	4.00E-05	0	-17.12
Muscone	Cinnamic acid	1.31E-03	8.43E-02	2.96	3.74E-05	0	-
Borneol	Bufalin	1.54E-03	6.82E-02	2.88	3.64E-05	0	-
Deoxycholic acid	Chenodeoxycholic acid	1.93E-03	5.11E-02	2.78	3.54E-05	0	-21.71
17-hydroxyprogesterone	11-hydroxyprogesterone	1.80E-03	5.51E-02	3.19	3.10E-05	0	-
Hyodeoxycholic acid	Chenodeoxycholic acid	1.92E-03	3.73E-02	2.85	2.52E-05	0	-19.83
Cinnamaldehyde	Benzyl benzoate	1.18E-03	6.27E-02	3.13	2.37E-05	0	-
Borneol	Cinnamaldehyde	1.32E-03	4.79E-02	2.97	2.13E-05	0	-
Benzyl benzoate	Cinnamic acid	1.10E-03	6.07E-02	3.17	2.11E-05	0	-
Borneol	Cinnamic acid	1.24E-03	4.87E-02	3.00	2.01E-05	0	-
Hyodeoxycholic acid	Deoxycholic acid	1.92E-03	2.78E-02	2.85	1.87E-05	0	-
Gamabufotalin	Bufalin	1.54E-03	1.79E-02	2.87	9.61E-06	0	-6.57
Cinnamaldehyde	Cinnamic acid	1.04E-03	1.44E-02	3.07	4.87E-06	0	-

4.3.8 Suggested modes-of-action of synergistic combinations from a target-network perspective

In this section, the MOAs of four synergistic combinations were suggested from the highly linked targets of the compounds.

The first combination that shows the highest SS ($SS=1.10E-03$) is ginsenoside Rb3 and cholic acid. Ginsenoside Rb3 was predicted to modulate four targets, in addition to three targets (Caspase-3 (CASP3) (301), Caspase-8 (CASP8) (301) and Caspase-9 (CASP9) (301)), which were extracted from literature. In total, 201 GO BP were annotated to these seven targets. Cholic acid was predicted to modulate 18 targets, which were annotated with 197 GO biological processes. The list of predicted targets of all combinations per discussion can be found in **Table 4.9**. The combination did not share any common predicted targets (**Figure 4.7a**). However, it was found that both compounds shared 143 common GO BP. Studies have been reported on the implications of some of the predicted targets and GO BP in cell proliferation. The highest degree node of cholic acid, androgen receptor (AR) ($k=206$) has been reported to modulate androgen-mediated vascular endothelial functions by interacting with the VEGF receptor signalling pathway upon androgen treatment (302). AR is also shown to recruit Src and activate both the MAPK pathway and the PI3K-Akt cascade leading to cell survival and proliferation (MAPK cascade, GO:0000165, $p\text{-value}=2.4E-165$) (303). The next highly connected predicted targets of cholic acid were glucocorticoid receptor (NR3C1) ($k=126$) (304), estrogen receptor (ESR1) ($k=45$) (305), nitric oxide synthase, inducible (NOS2) ($k=38$) and tyrosine-protein phosphatase non-receptor type 1 (PTPN1) ($k=26$) (306), which have been reported for their activities in vascular angiogenesis. Fibroblast growth factor 1 (FGF1) ($k=41$), the highest degree node of ginsenoside Rb3, is involved in tissue repair and regeneration, with FGF-induced signalling mainly mediated via the MAPK pathway (307). The Bliss score is 6.51. Therefore, the synergy of ginsenoside Rb3 and cholic acid is suggested to arise from the modulation of different targets of related pathways.

Combination of ginsenoside Rb3 and ginsenoside Rb2 shows an SS of $9.97E-04$, with both compounds predicted to modulate four identical targets (**Figure 4.7b**). From literature, three additional targets were associated to the MOA of ginsenoside Rb3 and four additional targets were associated to the MOA of ginsenoside Rb2. 195 GO BP were linked to targets of ginsenoside Rb3, while 201 GO BP were associated with ginsenoside Rb2. The combinations shared 163 common GO BP. The four additional targets of ginsenoside Rb2 were epidermal growth factor receptor (EGFR) ($k=211$), fibronectin (FN1) ($k=143$), pro-epidermal growth factor (EGF) ($k=108$) and interstitial collagenase (MMP1) ($k=32$), which were extracted from a study of the wound healing effects of ginsenoside Rb2 (308). CASP3 ($k=27$), CASP8 ($k=17$)

and CASP9 ($k=5$) were also extracted from literature. Ginsenoside Rb3 has been shown to increase cell viability in ischemic model by significantly attenuated the increase of caspase activities in a dose-dependent manner (301). In addition, FGF1 ($k=41$), one of the predicted targets of ginsenoside Rb3, also promotes cell proliferation (307). EGFR, EGF, and FN1 were annotated with regulation of angiogenesis (GO:0045765, $p\text{-value}=5.5\text{E-}51$), while the caspase targets were annotated with positive regulation of cell death (GO:0010942, $p\text{-value}=1.6\text{E-}103$). The inverse actions by two groups of targets might contribute to the large Bliss score of 13.99, despite the combinations sharing many common GO BPs. Therefore, the combination of ginsenoside Rb3 and ginsenoside Rb2 is proposed to exhibit synergy through modulating different targets at unrelated pathways.

The third combination, ginsenoside Rb3 and 11-hydroxyprogesterone ($SS=9.09\text{E-}04$) shows two sets of different predicted targets (**Figure 4.7c**). 11-hydroxyprogesterone was predicted to modulate 39 targets that were associated with 274 GO BP. 201 GO BP were associated with ginsenoside Rb3. The combination shared 185 common GO BP. In addition to AR in the list of 39 predicted targets of 11-hydroxyprogesterone and some of the highly linked targets such as AR ($k=206$), retinoic acid receptor RXR-alpha (RXRA) ($k=157$) (309), hypoxia-inducible factor 1-alpha (HIF1A) ($k=124$) (310), estrogen receptor (ESR1) ($k=45$) (305), nuclear factor erythroid 2-related factor 2 (NFE2L2) ($k=20$) (311) are implicated in cardiac vascular angiogenesis. Similar to the previous discussion, FGF1 of ginsenoside Rb3 is implicated in cell proliferation (307) and the inhibition of CASP3, CASP8, and CASP9 has been reported to induce angiogenesis (301). FGF1, HIF1A, and NFE2L2 were annotated with angiogenesis (GO:0001525, $p\text{-value}=2.2\text{E-}91$) and blood vessel development (GO:0001568, $p\text{-value}=2\text{E-}111$). The Bliss score of the combination is 7.67. Thus, the synergy between ginsenoside Rb3 and 11-hydroxyprogestrones is believed to be from the modulation of different targets of related pathways.

The fourth combination from **Table 4.7** that shows synergy is between ginsenoside Rd and ginsenoside Rb2 ($SS=8.60\text{-}4$) (**Figure 4.7d**). Ginsenoside Rd was predicted to modulate 12 targets that were annotated to 288 GO BP, of which 183 GO BP were common for both compounds in the combination. The combinations shared two common predicted targets, heparanase (HPSE) ($k=4$) and GBA ($k=1$), which might not play significant roles in vasculogenesis from the lower k value, but one common highly linked node, FGF1 ($k=41$). A few of the highly linked predicted targets of ginsenoside Rd are associated to vascular angiogenesis such as FGF1 ($k=41$), tyrosine-protein phosphatase non-receptor type 1 (PTPN1) ($k=46$) (306), SIRT1 ($k=25$) (312), apoptosis regulator Bcl-2 (BCL2) ($k=22$) (313) and apoptosis regulator BAX (BAX) ($k=19$). The association of ginsenoside Rb2's predicted

targets with cell proliferation have been discussed previously. Many of the highly connected predicted targets of both were compounds sharing common GO biological processes such as angiogenesis (GO:0001525, p-value=2.2E-91), blood vessel morphogenesis (GO:0048514, p-value=2.6E-100) and cell migration (GO:001647, p-value= 6.7E-133). The Bliss score of the combination is -3.71, which is still higher compare to the SS values of combinations in the bottom 20. Therefore, the synergy of the combination is suggested from the modulation of different targets of related pathways.

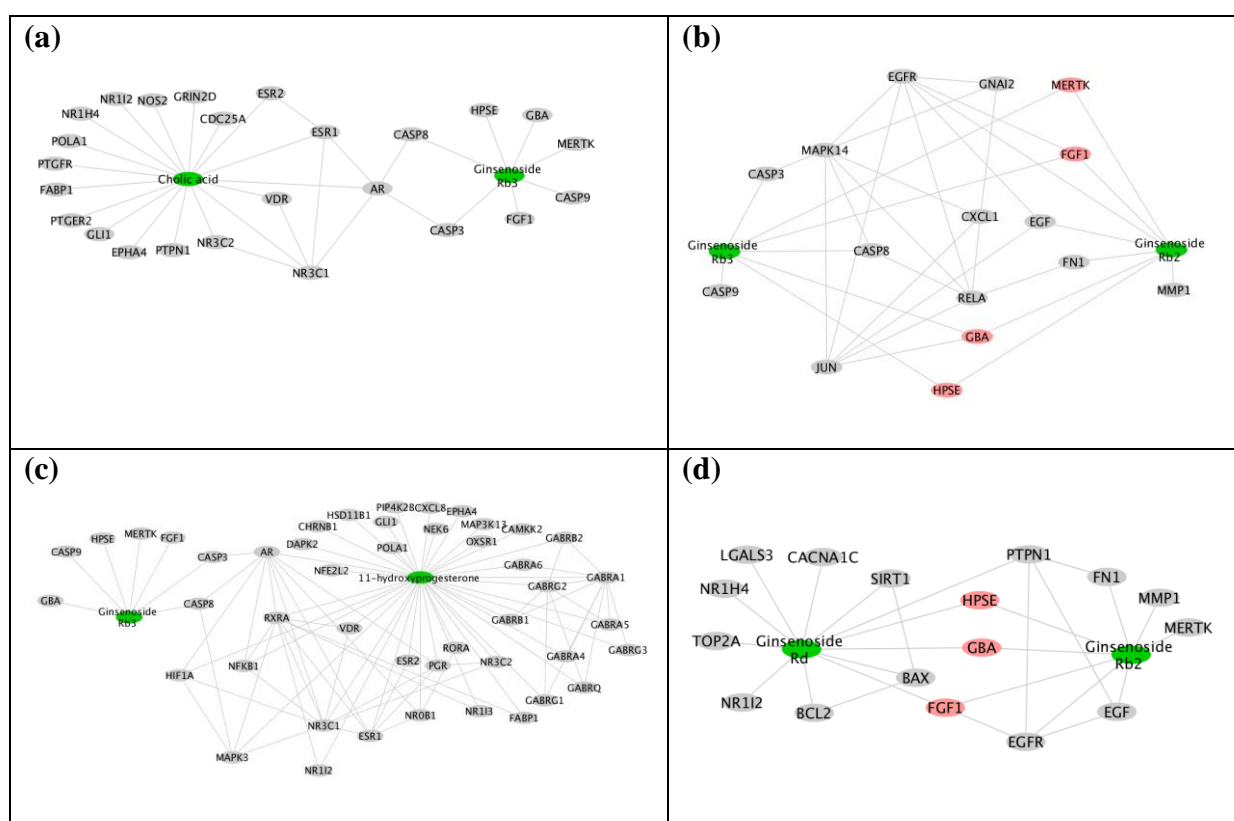


Figure 4.7: The networks of predicted synergistic combinations, which also show synergistic effects *in vitro*. Each network is composed of compounds that are linked to their predicted targets. In general, the predicted targets of both compounds (green circle) are observed to be linked by at least one target from one of the compounds (such in (a)) or shared predicted targets (both (b) and (d)) (red circle).

Table 4.9: The list of predicted targets of four combinations with measured Bliss synergy values from the top-ranked group.

Compound <i>i</i>	Compound <i>j</i>	Predicted target <i>i</i>	Predicted target <i>j</i>
Ginsenoside Rb3	Cholic acid	Fibroblast growth factor 1 (FGF1)	Androgen receptor (AR)
		Tyrosine-protein kinase Mer (MERTK)	Glucocorticoid receptor (NR3C1)
		Heparanase (HPSE)	Estrogen receptor (ESR1)
		Glucosylceramidase (GBA)	Nitric oxide synthase, inducible (NOS2)
		Caspase-3 (CASP3)	Tyrosine-protein phosphatase non-receptor type 1 (PTPN1)
		Caspase-8 (CASP8)	Estrogen receptor beta (ESR2)
		Caspase-9 (CASP9)	Ephrin type-A receptor 4 (EPHA4)
			Zinc finger protein GLI1 (GLI1)
			Mineralocorticoid receptor (NR3C2)
			Vitamin D3 receptor (VDR)
			Glutamate receptor ionotropic, NMDA 2D (GRIN2D)
			Nuclear receptor subfamily 1 group I member 2 (NR1I2)
			M-phase inducer phosphatase 1 (CDC25)
			Bile acid receptor (NR1H4)
			Fatty acid-binding protein, liver (FABP1)
			DNA polymerase alpha catalytic subunit (POLA1)
			Prostaglandin F2-alpha receptor (PTGFR)
			Prostaglandin E2 receptor EP4 subtype (PTGER4)
			Prostaglandin E2 receptor EP2 subtype (PTGER2)
Ginsenoside Rb3	Ginsenoside Rb2	Fibroblast growth factor 1 (FGF1)	Epidermal growth factor receptor (EGFR)
		Tyrosine-protein kinase Mer (MERTK)	Fibronectin (FN1)
		Glucosylceramidase (GBA)	Pro-epidermal growth factor (EGF)
		Caspase-3 (CASP3)	Fibroblast growth factor 1 (FGF1)
		Caspase-8 (CASP8)	Interstitial collagenase (MMP1)
		Caspase-9 (CASP9)	Tyrosine-protein kinase Mer (MERTK)
			Heparanase (HPSE)
			Glucosylceramidase (GBA)
Ginsenoside Rb3	11-hydroxyprogesterone	Fibroblast growth factor 1 (FGF1)	Androgen receptor (AR)
		Tyrosine-protein kinase Mer (MERTK)	Retinoic acid receptor RXR-alpha (RXRA)
		Glucosylceramidase (GBA)	Glucocorticoid receptor (NR3C1)
		Caspase-3 (CASP3)	Hypoxia-inducible factor 1-alpha (HIF1A)

Caspase-8 (CASP8)

Caspase-9 (CASP9)

Mitogen-activated protein kinase 3 (MAPK3)

Nuclear factor NF-kappa-B p105 subunit (NFKB1)

Estrogen receptor (ESR1)

Nuclear factor erythroid 2-related factor 2 (NFE2L2)

Gamma-aminobutyric acid receptor subunit alpha-1 (GABRA1)

Interleukin-8 (CXCL8)

Estrogen receptor beta (ESR2)

Progesterone receptor (PGR)

Ephrin type-A receptor 4 (EPHA4)

Zinc finger protein GLI1 (GLI1)

Gamma-aminobutyric acid receptor subunit gamma-2 (GABRG2)

Phosphatidylinositol 5-phosphate 4-kinase type-2 beta (PIP4K2B)

Mineralocorticoid receptor (NR3C2)

Vitamin D3 receptor (VDR)

Nuclear receptor ROR-alpha (RORA)

Nuclear receptor subfamily 0 group B member 1 (NR0B1)

Nuclear receptor subfamily 1 group I member 2 (NR1I2)

Gamma-aminobutyric acid receptor subunit beta-1 (GABRB1)

Mitogen-activated protein kinase kinase kinase 13 (MAP3K13)

Serine/threonine-protein kinase OSR1 (OSR1)

Fatty acid-binding protein, liver (FABP1)

Gamma-aminobutyric acid receptor subunit alpha-5 (GABRA5)

Gamma-aminobutyric acid receptor subunit beta-2 (GABRB2)

Nuclear receptor subfamily 1 group I member 3 (NR1I3)

Gamma-aminobutyric acid receptor subunit gamma-1 (GABRG1)

Calcium/calmodulin-dependent protein kinase kinase 2 (CAMKK2)

Serine/threonine-protein kinase Nek6 (NEK6)

Gamma-aminobutyric acid receptor subunit theta (GABRQ)

Gamma-aminobutyric acid receptor subunit gamma-3 (GABRG3)

Death-associated protein kinase 2 (DAPK2)

DNA polymerase alpha catalytic subunit (POLA1)

Acetylcholine receptor subunit beta (CHRNA1)

Corticosteroid 11-beta-dehydrogenase isozyme 1 (HSD11B1)

Gamma-aminobutyric acid receptor subunit alpha-4 (GABRA4)

Ginsenoside Rd	Ginsenoside Rb2	<p>Fibroblast growth factor 1 (FGF1)</p> <p>Tyrosine-protein phosphatase non-receptor type 1 (PTPN1)</p> <p>NAD-dependent protein deacetylase sirtuin-1 (SIRT1)</p> <p>Apoptosis regulator Bcl-2 (BCL2)</p> <p>Apoptosis regulator BAX (BAX)</p> <p>Voltage-dependent L-type calcium channel subunit alpha-1C (CACNA1C)</p> <p>DNA topoisomerase 2-alpha (TOP2A)</p> <p>Heparanase (HPSE)</p> <p>Nuclear receptor subfamily 1 group I member 2 (NR1I2)</p> <p>Bile acid receptor (NR1H4)</p> <p>Glucosylceramidase (GBA)</p> <p>Galectin-3 (LGALS3)</p>	<p>Epidermal growth factor receptor (EGFR)</p> <p>Fibronectin (FN1)</p> <p>Pro-epidermal growth factor (EGF)</p> <p>Fibroblast growth factor 1 (FGF1)</p> <p>Interstitial collagenase (MMP1)</p> <p>Tyrosine-protein kinase Mer (MERTK)</p> <p>Heparanase (HPSE)</p> <p>Glucosylceramidase (GBA)</p>
----------------	-----------------	---	--

4.3.9 Discussion on the suggested MOA of synergistic combinations

To this end, four combinations with SS values agree with Bliss scores measured from the response of cell growths of HUVEC. The synergistic effects of these combinations are suggested from the highly connected predicted targets of both compounds. In general, three combinations are suggested to exhibit synergistic effects by modulating different targets of related pathways and one combination is estimated to exert synergy by modulating different targets of unrelated pathways. The proposed types of MOA of compound combinations have been reported before; synergistic drug combinations can exert their activities by modulating similar/different targets of similar/different pathways (314). For ginsenoside Rb3, the synergistic MOA was proposed by modulating FGF1 and inhibiting the activation of caspases. Ginsenoside Rb2 was estimated to exhibit synergy from the modulation of growth factor related targets such as EGFR and EGF. Both cholic acid and 11-hydroxprogesterone were proposed to induce synergy from modulating AR. From the four synergistic combinations, the combination that contains either ginsenoside Rb2 and ginsenoside Rb3 was discovered to promote cell proliferation when they are combined with each other or 11-hydroxyprogesterone, cholic acid and ginsenoside Rd. The synergistic effect of ginsenoside Rb3 and ginsenoside Rb2 combination seems to be suggested from literature rather than the predicted targets but the information from such studies is not always available for other components of SBP. For example, many studies have been conducted independently relating to ginsenoside Rb1 and ginsenoside Re (315, 316), hence the network-based approach could provide an opportunity to identify the synergy between these compounds from a protein-protein interaction perspective.

In this study, some of the highly connected predicted targets such as mitogen-activated protein kinase 3 (MAPK3) ($k=77$) and protein kinase C eta type (PRKCH) ($k=20$) have a limited number of studies to link their associations with angiogenesis. In a few cases, low connected predicted targets, such as galectin-3 (LGALS3) ($k=1$) (317) and HPSE ($k=4$) (318) have been implicated in vascular angiogenesis implying the targets might be important players in a real biological network but the significance of these predicted targets are not captured in the network. The network is only a small representation of a real biological network, where a lot of information is not included (319). However, the combinations of the bottom rank group, which have lower SS values correspond to the Bliss synergy scores, suggest the network is still robust to the set of predicted targets of the compound combination.

Despite the proposed MOAs of the combinations from the top-ranked group, one combination in the group did not agree with the experimental results. The combination, ginsenoside Rb2 and ginsenoside Rg1 with SS value of 8.64E-04 was determined to show no

synergy based on Bliss score. The measured Bliss score was -10.29. From the single dose response curve in Appendices (**Figure A4.1**), ginsenoside Rg1 has already inhibited cell growth at a low concentration. The observation disagrees with a study that reports ginsenoside Rg1 increases the interactions between glucocorticoid receptor (NR3C1) and FGF1, which activate PI3K/Akt/eNOS signalling and eventually enhancing angiogenesis (320). Both of these targets are among the highly connected predicted targets of both compounds. However, the validation can be further confirmed by repeating the experiment. Therefore, to confirm the MOAs of the compounds by modulating the predicted targets, it is necessary to perform subsequent experiment such as Western blotting to detect expression of the predicted targets.

4.4 Conclusion

In this work, the evaluation of the synergy of 231 pairs of combinations from 22 SBP were measured from a representative network of angiogenesis and CHD. The representative network of angiogenesis and CHD was constructed using a knowledge-based approach which is demonstrated to be feasible in estimating synergistic SBP compound combinations. In the first part of the study, the representative network was built by integrating different data sources to identify genes annotated with angiogenesis and CHD. By utilising information of PPI extracted from Reactome and Signalink2.0, the network was shown to be a scale free network and represented the “desired” biological context where the top ten hub genes were related to the biological process of angiogenesis/CHD. In the second part of the study, three elements were formulated into an equation to measure synergy of two-compound combinations. The three elements were topological properties, pathway dissimilarity and mean distance of all the predicted targets of a combination. The integration of the three elements are reasonable to measure how the predicted targets are linked to each other in the network with their related pathways to produce synergy. The model was robust when compared to the background distribution of random networks with permuted edges and random targets. In the third part of the study, the results from experimental validation indicate four combinations from the 20 top-ranked combinations agree with the SS. These combinations are ginsenoside Rb3 and cholic acid, ginsenoside Rb2 and ginsenoside Rb3, ginsenoside Rb3 and 11-hydroxyprogesterone and ginsenoside Rb2 and ginsenoside Rd. The Bliss scores were higher than the scores of five combinations in the bottom ranks. The MOA of the three synergistic combinations were proposed from the modulation of predicted targets of related pathways. One synergistic combination showed synergy from modulating the different targets of unrelated pathways. Despite one combination in the top-ranked group disagreeing with experimental result, the approach of utilising systems biology provides the opportunity to identify simplified combinations with a desirable “network” effect.

Chapter 5:

Concluding Remarks

This thesis concentrated on understanding the MOA of TM compounds by applying an *in silico* target prediction. A chemogenomic database of ligand-target pairs is developed from bioactivity databases that form a training set of the model. The model learns the pattern of chemical spaces of the ligands to their protein targets via two machine learning models, Naïve Bayes Classifier and Random Forest. To predict a target class of a new compound, the model compares the structural similarity of the new compound to those compounds in the model based on the “Molecular Similarity Principle”. Hence, the *in silico* target prediction has the ability to correlate the chemical descriptor of TM compounds to their protein targets, which the MOA can be rationalise from associating the indications of the TM compounds (plants) to their predicted targets.

In Chapter 2, we showed that the applicability of an *in silico target* prediction improved mechanistic understanding of the relationship between 46 TCM therapeutic action classes and subclasses, which are not apparent from the name given to the (sub-)class itself, via the dendrogram of 14 clusters. Despite different names are given to the subclasses, the frequent enriched predicted targets and annotated pathways are related to immune related proteins and digestive system respectively. The findings could be linked to the view of TCM, where symptoms are usually regarded as the invasion of pathogenic factors, thus sensitizing the immune system to response and the important treatment principle is “the foundation of acquired constitution” that includes spleen and stomach. The results from the study can be further developed by studying the connections between predicted targets from the significant clusters of protein family and pathway motif. This study could help to better understand the MOAs of TCM compounds in maintaining a healthy body system by relating the digestive system to immune-related targets of the GPCR or protein kinase family.

In Chapter 3, we explored the similarity and differences of the chemical space and bioactivity space of TCM, Ayurveda and Malay TM for treating cancer. The key findings indicated that they shared six frequent scaffolds and protein kinase was the most significant target class in all three TMs. The small similarity suggested that many of the remaining compounds are worth exploring for their anti-cancer activities, as additional nine frequent scaffolds are found to abundance in the anti-cancer plants and the compounds were predicted to modulate additional 13 targets classes. The comparison of the bioactivity space discovered the majority of the plants were not phylogenetically related except for eight pairs of plants. The findings of the large compound diversity and high similarity of MOAs could lead to new

knowledge of the cancer pathogenesis by identifying which biological pathways these targets are annotated with. It is suggested that the chemical and bioactivity space exploration strategies could be a stepping-stone to compare more TM datasets of different countries. The study could provide an opportunity discover more compound diversity for treating cancer.

In Chapter 4, we evaluated the synergy of pairwise compound combinations of 22 SBP compounds by extending the target prediction and pathway annotation to network topology in order to understand the MOA of SBP for CHD by promoting angiogenesis. The integration of three parameters, topology, pathway dissimilarity and mean distance of the set of predicted targets of a compound combination to measure synergy is reasonable to measure how the predicted targets are linked to each other in the network with their related pathways. The findings of four synergistic combinations that agreed with experimental results demonstrate that the method can be potentially applied to explore compound combinations of other TCM formulations with limited bioactivity data to start with. To confirm the predicted targets are responsible for the MOAs of the synergistic combinations, further experimental studies such as gene expression analysis and Western blotting are proposed.

Although the studies have successfully explained the link of predicted targets and pathways in suggesting the MOAs of TM compounds, the studies are still limited to targets that are only available in the *in silico* target prediction. For example, the protein kinase family is the biggest group of protein family in the model. Thus, extending protein targets of other protein families could provide a more comprehensive overview of targets that are implicated for the indications of the TM compounds. The available chemical space of the model is limited to the version of ChEMBL used when the model was developed. In addition, natural compounds only represent approximately 3.85% of the total compounds available. Therefore, the limited coverage of chemical space leads to no predicted targets for one of the TCM therapeutic action subclasses, “tranquilizing, settling” in Chapter 1. Despite the limitations, our *in silico* target prediction was able to describe a reasonable MOA of the TM compounds, hence, the analyses could be beneficial in providing testable hypotheses to guide towards finding new molecular entities.

References:

1. M. Lahlou, The Success of Natural Products in Drug Discovery. *Pharmacology & Pharmacy* **4**, 17-31 (2013).
2. D. J. Newman, G. M. Cragg and K. M. Snader, The influence of natural products upon drug discovery. *Natural Product Reports* **17**, 215-234 (2000).
3. R. S. Solecki, Shanidar IV, a Neanderthal Flower Burial in Northern Iraq. *Science* **190**, 880-881 (1975).
4. D. A. Dias, S. Urban and U. Roessner, A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2**, 303-336 (2012).
5. H.-F. Ji, X.-J. Li and H.-Y. Zhang, Natural products and drug discovery. *EMBO reports* **10**, 194-200 (2009).
6. *Traditional, complementary and integrative medicine* (<http://www.who.int/traditional-complementary-integrative-medicine/about/en/>), (2017).
7. C. Fisch, William Withering: An Account of the Foxglove and Some of its Medical Uses. *JACC* **5**, 1A-2A (1985).
8. G. M. Rishton, Natural Products as a Robust Source of New Drugs and Drug Leads: Past Successes and Present Day Issues *The American Journal of Cardiology* **101**, S43-S49 (2008).
9. D. S. Fabricant and N. R. Farnsworth, The Value of Plants Used in Traditional Medicine for Drug Discovery. *Environmental and Health Perspectives* **109**, 69-75 (2001).
10. D. J. Newman and G. M. Cragg, Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products* **79**, 629-661 (2016).
11. A. Ganesan, The impact of natural products upon modern drug discovery. *Current Opinion in Chemical Biology* **12**, 306-317 (2008).
12. D. G. I. Kingston, Modern Natural Products Drug Discovery and its Relevance to Biodiversity Conservation. *Journal of Natural Products* **74**, 496-511 (2011).
13. W. R. Strohl, The role of natural products in modern drug discovery. *Drug Discovery Today* **5**, 39-41 (2000).
14. M. S. Butler, The Role of Natural Product Chemistry in Drug Discovery. *Journal of Natural Products* **67**, 2141-2153 (2004).
15. U. S. Eggert, The why and how of phenotypic small-molecule screens. *Nature Chemical Biology* **9**, 206-209 (2013).
16. D. C. Swinney and J. Anthony, How were new medicines discovered? *Nature Reviews Drug Discovery* **10**, 507-519 (2011).
17. M. Schenone, V. Dancik, B. K. Wagner and P. A. Clemons, Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology* **9**, 232-240 (2013).
18. J. N. Y. Chan, C. Nislow and A. Emili, Recent advances and method development for drug target identification. *Trends in Pharmacological Sciences* **31**, 82-88 (2010).
19. A. D. B. Vaidya, Reverse Pharmacology - A Paradigm Shift for Drug Discovery and Development. *Current Research in Drug Discovery* **1**, 39-44 (2014).
20. J. Xu and A. Hagler, Chemoinformatics and Drug Discovery. *Molecules* **7**, 566-600 (2002).
21. P. Willett, J. M. Barnard and G. M. Downs, Chemical Similarity Searching. *Journal of Chemical Information and Computer Sciences* **38**, 983-996 (1998).
22. J. L. Medina-Franco, Interrogating Novel Areas of Chemical Space for Drug Discovery using Chemoinformatics. *Drug Development Research* **73**, 430-438 (2012).
23. D. Rognan, Chemogenomic approaches to rational drug design. *British Journal of Pharmacology* **152**, 38-52 (2007).
24. M. Bredel and E. Jacoby, Chemogenomics: An emerging strategy for rapid target and drug discovery. *Nature Reviews Genetics* **5**, 262-275 (2004).
25. H. E. Pence and A. Williams, ChemSpider: An Online Chemical Information Resource. *Journal of Chemical Education* **87**, 1123-1124 (2010).
26. T. Sterling and J. J. Irwin, ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling* **55**, 2324-2337 (2015).
27. M. K. Gilson, T. Liu, M. Baitaluk *et al.*, BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research* **44**, D1045-D1053 (2016).

28. A. Gaulton, L. J. Bellis, A. P. Bento *et al.*, ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* **40**, D1100-D1107 (2012).
29. S. Kim, P. A. Thiessen, E. E. Bolton *et al.*, PubChem Substance and Compound databases. *Nucleic Acids Research* **44**, D1202-D1213 (2016).
30. D. S. Wishart, C. Knox, A. C. Guo *et al.*, DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research* **36**, D901-D906 (2008).
31. F. Zhu, Z. Shi, C. Qin *et al.*, Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Research* **40**, D1128-D1136 (2012).
32. J. Ru, P. Li, J. Wang *et al.*, TCMSP: a database of systems pharmacology from herbal medicines. *Journal of Cheminformatics* **6**, 1-6 (2014).
33. M. Mangal, P. Sagar, H. Singh, G. P. S. Raghava and S. M. Agarwal, NPACT: Naturally Occurring Plant-based Anti-cancer Compound-Activity-Target database. *Nucleic Acids Research* **41**, D1124-D1129 (2013).
34. P. Banerjee, J. Erehman, B.-O. Gohlke *et al.*, Super Natural II—a database of natural products. *Nucleic Acids Research* **43**, D935-D939 (2015).
35. E. v. d. Horst, J. E. Peironcelly, G. J. P. v. Westen *et al.*, Chemogenomics Approaches for Receptor Deorphanization and Extensions of the Chemogenomics Concept to Phenotypic Space. *Current Topics in Medicinal Chemistry* **11**, 1-14 (2011).
36. C. R. Andersson, M. G. Gustafsson and H. Strömbergsson, Quantitative Chemogenomics: Machine-Learning Models of Protein-Ligand Interaction. *Current Topics in Medicinal Chemistry* **11**, 1978-1993 (2011).
37. N. P. Savchuk, K. V. Balakin and S. E. Tkachenko, Exploring the chemogenomic knowledge space with annotated chemical libraries. *Current Opinion in Chemical Biology* **8**, 412-417 (2004).
38. E. Jacoby, Computational chemogenomics. *WIREs Computational Molecular Science* **1**, 57-67 (2011).
39. H. Strömbergsson and G. J. Kleywegt, A chemogenomics view on protein-ligand spaces. *BMC Bioinformatics* **10**, 1-11 (2009).
40. A. R. Leach and V. J. Gillet, *An Introduction to Chemoinformatics*. (Springer, The Netherlands, 2007), pp. 255.
41. A. Bender and R. C. Glen, Molecular similarity: a key technique in molecular informatics. *Organic Biomolecular Chemistry* **2**, 3204-3218 (2004).
42. J. B. O. Mitchell, Machine learning methods in chemoinformatics. *WIREs Computational Molecular Science* **4**, 468-481 (2014).
43. V. Poroikov, D. Filimonov, A. Lagunin, T. Glorizova and A. Zakharov, PASS: identification of probable targets and mechanisms of toxicity. *SAR and QSAR in Environmental Research* **18**, 101-110 (2007).
44. Nidhi, M. Glick, J. W. Davies and J. L. Jenkins, Prediction of Biological Targets for Compounds Using Multiple-Category Bayesian Models Trained on Chemogenomics Databases. *Journal of Chemical Information and Modeling* **46**, 1124-1133 (2006).
45. M. J. Keiser, B. L. Roth, B. N. Armbruster *et al.*, Relating protein pharmacology by ligand chemistry. *Nature Biotechnology* **25**, 197-206 (2007).
46. J. Mestres, L. Martín-Couce, E. Gregori-Puigjané, M. Cases and S. Boyer, Ligand-Based Approach to *In Silico* Pharmacology: Nuclear Receptor Profiling. *Journal of Chemical Information and Modeling* **46**, 2725-2736 (2006).
47. F. M. Fauzi, A. Koutsoukas, R. Lowe *et al.*, Chemogenomics Approaches to Rationalizing the Mode-of-Action of Traditional Chinese and Ayurvedic Medicines. *Journal of Chemical Information and Modeling* **53**, 661-673 (2013).
48. D. Rogers and M. Hahn, Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **50**, 742-754 (2010).
49. A. Bender, D. W. Young, J. L. Jenkins *et al.*, Chemogenomic Data Analysis: Prediction of Small-Molecule Targets and the Advent of Biological Fingerprints. *Combinatorial Chemistry & High Throughput Screening* **10**, 719-731 (2007).
50. X. Chen, C. Y. Ung and Y. Chen, Can an *in silico* drug-target search method be used to probe potential mechanisms of medicinal plant ingredients? *Natural Product Reports* **20**, 432-444 (2003).

51. S. Zhang, W. Lu, X. Liu *et al.*, Fast and effective identification of the bioactive compounds and their targets from medicinal plants via computational chemical biology approach. *Medicinal Chemistry Communication* **2**, 471-477 (2011).
52. J. L. Jenkins, A. Bender and J. W. Davies, *In silico* target fishing: Predicting biological targets from chemical structure. *Drug Discovery Today: Technologies* **3**, 413-421 (2006).
53. A. E. Cleves and A. N. Jain, Robust Ligand-Based Modeling of the Biological Targets of Known Drugs. *Journal of Medicinal Chemistry* **49**, 2921-2938 (2005).
54. T. Cheng, Q. Li, Y. Wang and S. H. Bryant, Identifying Compound-Target Associations by Combining Bioactivity Profile Similarity Search and Public Databases Mining. *Journal of Chemical Information and Modeling* **51**, 2440-2448 (2011).
55. H. Kitano, Systems Biology: A Brief Overview. *Science* **295**, 1662-1664 (2002).
56. E. C. Butcher, Can cell systems biology rescue drug discovery? *Nature Reviews Drug Discovery* **4**, 461-467 (2005).
57. G. Apic, T. Ignjatovic, S. Boyer and R. B. Russell, Illuminating drug discovery with biological pathways. *FEBS Letters* **579**, 1872-1877 (2005).
58. A. Pujol, R. Mosca, J. Farrés and P. Aloy, Unveiling the role of network and systems biology in drug discovery. *Trends in Pharmacological Sciences* **31**, 115-123 (2010).
59. M. E. Cusick, H. Hu and A. Smolyar, Literature-curated protein interaction datasets. *Nature Methods* **6**, 39-46 (2009).
60. T. Klingström and D. Plewczynski, Protein-protein interaction and pathway databases, a graphical review. *Briefings in Bioinformatics* **12**, 702-713 (2010).
61. H. Hermjakob, L. Montecchi-Palazzi and C. Lewington, IntAct: An open source molecular interaction database. *Nucleic Acids Research* **32**, D452-D455 (2004).
62. T. S. Keshava Prasad, R. Goel, K. Kandasamy *et al.*, Human Protein Reference Database—2009 update. *Nucleic Acids Research* **37**, D767-D772 (2009).
63. C. Stark, B. J. Breitkreutz and T. Reguly, BioGRID: A general repository for interaction datasets. *Nucleic Acids Research* **34**, D535-D539 (2006).
64. A. Zanzoni, L. Montecchi-Palazzi and M. Quondam, MINT: A Molecular INTERaction database. *FEBS Letters* **513**, 135-140 (2002).
65. M. Kanehisa and S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27-30 (2000).
66. D. Fazekas, M. Koltai, D. Türei *et al.*, Signalink 2 – a signaling pathway resource with multi-layered regulatory networks. *BMC Systems Biology* **7**, 1-7 (2013).
67. G. Joshi-Tope, M. Gillespie, I. Vastrik *et al.*, Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research* **33**, D428-D432 (2005).
68. S. Liggi, G. Drakakis, A. Henry *et al.*, Extensions to *In Silico* Bioactivity Predictions Using Pathway Annotations and Differential Pharmacology Analysis: Application to *Xenopus laevis* Phenotypic Readouts. *Journal of Molecular Informatics* **32**, 1009-1024 (2013).
69. J. Liu, K. Sun, C. Zheng *et al.*, Pathway as a Pharmacological Target for Herbal Medicines: An Investigation from *Reduning Injection*. *PLoS ONE* **10**, e0123109 (2015).
70. N. Le Novère, Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics* **16**, 146-158 (2015).
71. K. Shahzad and J. J. Loor, Application of Top-Down and Bottom-up Systems Approaches in Ruminant Physiology and Metabolism. *Current Genomics* **13**, 379-394 (2012).
72. R. Liu, C.-X. Guo and H.-H. Zhou, Network-based approach to identify prognostic biomarkers for estrogen receptor-positive breast cancer treatment with tamoxifen. *Cancer Biology & Therapy* **16**, 317-324 (2015).
73. W. Zhu, L. Yang and Z. Du, Layered Functional Network Analysis of Gene Expression in Human Heart Failure. *PLoS ONE* **4**, e6288 (2009).
74. G. M. Maggiora, On Outliers and Activity Cliffs Why QSAR Often Disappoints. *Journal of Chemical Information and Modeling* **46**, 1535 (2006).
75. T. Kalliokoski, C. Kramer, A. Vulpetti and P. Gedeck, Comparability of Mixed IC₅₀ Data – A Statistical Analysis. *PLoS ONE* **8**, e61007 (2013).
76. K. Chan, Progress in traditional Chinese medicine. *Trends in Pharmacological Sciences* **16**, 182-188 (1995).
77. J. B. Waldram, The Efficacy of Traditional Medicine: Current Theoretical and Methodological Issues. *Medical Anthropology* **14**, 603-605 (2008).

78. K. Chen and H. Xu, The integration of traditional Chinese medicine and Western medicine. *European Review* **11**, 225-235 (2003).
79. E. Chan, M. Tan, J. Xin, S. Sudarsanam and D. E. Johnson, Interactions between traditional Chinese medicines and Western therapeutics. *Current Opinion Drug Discovery & Development* **13**, 50-65 (2010).
80. Y. Wang and A. Xu, A systems biology approach to diagnosis and treatments. *Science* **346**, S13-S15 (2014).
81. X. Zhao, X. Zheng, T.-P. Fan *et al.*, A novel drug discovery strategy inspired by traditional medicine philosophies. *Science* **347**, S38-S40 (2015).
82. R. Liu, R. Runyon, Y. Wang *et al.*, Deciphering ancient combinatorial formula-The Shexiang Baixin pill. *Science* **347**, S40-S42 (2015).
83. W. Lam, S.-H. Liu, Z. Jiang and Y.-C. Cheng, Lessons from the development of the traditional Chinese medicine formula PHY906. *Science* **347**, S43-S44 (2015).
84. H. Sheriden, B. Kopp, L. Krenn, D. Guo and J. Sendker, Traditional Chinese herbal medicine preparation: Invoking the butterfly effect. *Science* **350**, S64-S66 (2015).
85. S. Li, Mapping ancient remedies: Applying a network approach to traditional Chinese medicine. *Science* **350**, S72-S74 (2015).
86. B. He, C. Lu, M. Wang *et al.*, Drug discovery in traditional Chinese medicine: From herbal fufang to combinatory drugs. *Science* **350**, S74-S76 (2015).
87. J. K. Chen and T. T. Chen, *Chinese Medical Herbalogy and Pharmacology*. L. Crampton, Ed., (Art of Medicine Press, Inc., City of Industry, California, 2004), pp. 1619.
88. C. Y. Chen, TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS ONE* **6**, e15939 (2011).
89. J. Zhao, P. Jiang and W. Zhang, Molecular networks for the study of TCM Pharmacology. *Briefings in Bioinformatics* **11**, 417-430 (2009).
90. M. Zhao, Q. Zhou, W. Ma and D.-Q. Wei, Exploring the Ligand-Protein Networks in Traditional Chinese Medicine: Current Databases, Methods, and Applications. *Evidence-Based Complementary and Alternative Medicine* **2013**, 1-15 (2013).
91. J. Qiu, A culture in the balance. *Nature* **448**, 126-128 (2007).
92. L. Wang, G.-B. Zhou, P. Liu *et al.*, Dissection of mechanisms of Chinese medicinal formula Realgar-Indigo naturalis as an effective treatment for promyelocytic leukemia. *Proceedings of the National Academy of Sciences* **105**, 4826-4831 (2008).
93. T.-P. Fan, J.-C. Yeh, K. W. Leung, P. Y. K. Yue and R. N. S. Wong, Angiogenesis: from plants to blood vessels. *Trends in Pharmacological Sciences* **27**, 297-309 (2006).
94. S. Ma, C. Feng, X. Zhang *et al.*, The multi-target capabilities of the compounds in a TCM used to treat sepsis and their *in silico* pharmacology. *Complementary Therapies in Medicine* **21**, 35-41 (2013).
95. Z. Xu, Modernization: One step at a time. *Nature Outlook* **480**, S90-S92 (2011).
96. T. W. Carson and C. M. Crews, Molecular Understanding and Modern Applications of Traditional Medicines: Triumphs and Trials. *Cell* **130**, 769-774 (2007).
97. Y. Cai, Q. Luo, M. Sun and H. Corke, Antioxidant activity and phenolic compounds of 112 traditional Chinese medicinal plants associated with anticancer. *Life Sciences* **74**, 2157-2184 (2004).
98. Q. Wang, H. Kuang, Y. Su *et al.*, Naturally derived anti-inflammatory compounds from Chinese medicinal plants. *Journal of Ethnopharmacology* **146**, 9-39 (2013).
99. Y. J. Kang, Herbogenomics: From Traditional Chinese Medicine to Novel Therapeutics. *Experimental Biology and Medicine* **2033**, 1059-1065 (2008).
100. S. Zahler, S. Tietze, F. Totzke *et al.*, Inverse *In Silico* Screening for Identification of Kinase Inhibitor Targets. *Chemistry & Biology* **14**, 1207-1214 (2007).
101. A. A. Lagunin, R. K. Goel, D. Y. Gawande *et al.*, Chemo- and bioinformatics resources for *in silico* drug discovery from medicinal plants beyond their traditional use: a critical review. *Natural Product Reports* **31**, 1585-1611 (2014).
102. Molecular Operating Environment (MOE), 2013.08, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, (2013).
103. A. Koutsoukas, R. Lowe, Y. Kalantar-Motamedi *et al.*, *In silico* target predictions: comparing multiclass Naïve Bayes and Parzen-Rosenblatt Window and the definition of a benchmarking

- dataset for target prediction. *Journal of Chemical Information and Modeling* **53**, 1957–1966 (2013).
104. A. Bender, H. Y. Mussa, R. C. Glen and S. Reiling, Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *Journal of Chemical Information and Computer Sciences* **44**, 1708-1718 (2004).
 105. P. A. Flach and N. Lachiche, Naive Bayesian Classification of Structured Data. *Machine Learning* **57**, 233-269 (2004).
 106. M. Olah, R. Rad, L. Ostopovici *et al.*, in *Chemical Biology*. (Wiley-VCH Verlag GmbH, 2007), pp. 760-786.
 107. G. Drakakis, A. Koutsoukas, S. C. Brewerton *et al.*, Comparing Global and Local Likelihood Score Thresholds in Multiclass Laplacian-Modified Naive Bayes Protein Target Prediction. *Combinatorial Chemistry & High Throughput Screening* **18**, 323-330 (2015).
 108. M. Kanehisa, S. Goto, Y. Sato *et al.*, Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* **42**, D199-D205 (2014).
 109. S. Liggi, G. Drakakis, A. Koutsoukas *et al.*, Extending *in silico* mechanism-of-action analysis by annotating targets with pathways: application to cellular cytotoxicity readouts. *Future Medicinal Chemistry* **6**, 2029-2056 (2014).
 110. E. E. Bolton, Y. Wang, P. A. Thiessen and S. H. Bryant, PubChem: Integrated Platform of Small Molecules and Biological Activities. Chapter 12 IN Wheeler RA and Spellmeyer DC, eds. *Annual Reports in Computational Chemistry* **4** 217-241 (2008).
 111. J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling* **52**, 1757-1768 (2012).
 112. W. Liu and D. E. Johnson, Clustering and its application in multi-target prediction. *Current Opinion Drug Discovery & Development* **12**, 98-107 (2009).
 113. R. D. Browne and Y. C. Martin, Structure-Activity Data Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *Journal of Chemical Information and Computer Sciences* **36**, 572-584 (1996).
 114. RStudio Team (2016). RStudio: Integrated development environment for R, RStudio, Inc., Boston, MA
 115. J. H. Ward, Hierarchical Grouping to Optimize an Objective Function. *Journal of American Statistical Association* **58**, 236-244 (1963).
 116. The UniProt Consortium, Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Research* **42**, D191-D198 (2014).
 117. P. Nussbaumer and A. Billich, Steroid Sulfatase Inhibitors. *Medicinal Research Reviews* **24**, 529-576 (2004).
 118. K. M. Doody, A. Bourdeau and M. L. Tremblay, T-cell protein tyrosine phosphatase is a key regulator in immune cell signaling: lessons from the knockout mouse model and implications in human disease. *Immunological Reviews* **228**, 325-341 (2009).
 119. A. M. Cameron, J. Frederick C. Nucifora, E. T. Fung *et al.*, FKBP12 Binds the Inositol 1,4,5-Trisphosphate Receptor at Leucine-Proline (1400–1401) and Anchors Calcineurin to this FK506-like Domain. *Journal of Biological Chemistry* **272**, 27582-27588 (1997).
 120. S. A. Jabbour and B. J. Goldstein, Sodium glucose co-transporter 2 inhibitors: blocking renal tubular reabsorption of glucose to improve glycaemic control in patients with diabetes. *International Journal of Clinical Practice* **62**, 1279-1284 (2008).
 121. J. L. Nitiss, Investigating the biological functions of DNA topoisomerases in eukaryotic cells. *Biochimica et Biophysica Acta* **1400**, 63-81 (1998).
 122. S. Andersson, W. M. Geissler, L. Wu *et al.*, Molecular genetics and pathophysiology of 17 beta-hydroxysteroid dehydrogenase 3 deficiency. *The Journal of Clinical Endocrinology & Metabolism* **81**, 130-136 (1996).
 123. C. Bařinka, C. Rojas, B. Slusher and M. Pomper, Glutamate Carboxypeptidase II in Diagnosis and Treatment of Neurologic Disorders and Prostate Cancer. *Current Medicinal Chemistry* **19**, 856-870 (2012).
 124. B. E. Shan, K. Zeki, T. Sugiura, Y. Yoshida and U. Yamashita, Chinese Medicinal Herb, *Acanthopanax gracilistylus*, Extract Induces Cell Cycle Arrest of Human Tumor Cells *in vitro*. *Japanese Journal of Cancer Research* **91**, 383-389 (2000).

125. L. Zhang, A. S. Ravipati, S. R. Koyyalamudi *et al.*, Antioxidant and Anti-inflammatory Activities of Selected Medicinal Plants Containing Phenolic and Flavonoid Compounds. *Journal of Agricultural and Food Chemistry* **59**, 12361-12367 (2011).
126. *Chinese Herbal Medicine Materia Medica 3rd Edition*. (Eastlan Press, Inc., Seattle, WA, 2004).
127. T. N. Hilton, L. J. Tuttle, K. L. Bohnert, M. J. Mueller and D. R. Sinacore, Excessive Adipose Tissue Infiltration in Skeletal Muscle in Individuals With Obesity, Diabetes Mellitus, and Peripheral Neuropathy: Association With Performance and Function *Physical Therapy* **88**, 1336-1344 (2008).
128. A. Giordano, M. Calvani, O. Petillo *et al.*, Skeletal Muscle Metabolism in Physiology and in Cancer Disease. *Journal of Cellular Biochemistry* **90**, 170-186 (2003).
129. P. Pouliot, S. Bergeron, A. Marette and M. Olivier, The role of protein tyrosine phosphatases in the regulation of allergic asthma: implication of TC-PTP and PTP-1B in the modulation of disease development. *Immunology* **128**, 534-542 (2009).
130. A. B. Hocaoglu, O. Karaman, D. O. Erge *et al.*, Glycyrrhizin and Long-Term Histopathologic Changes in a Murine Model of Asthma. *Current Therapeutic Research, Clinical and Experimental* **72**, 250-261 (2011).
131. J. H. Choi, Y. P. Hwang, H. G. Kim *et al.*, Saponins from the Roots of *Platycodon grandiflorum* Suppresses TGF β 1-Induced Epithelial-Mesenchymal Transition Via Repression of PI3K/Akt, ERK1/2 and Smad2/3 Pathway in Human Lung Carcinoma A549 Cells. *Nutrition and Cancer* **66**, 140-151 (2013).
132. S. R. Kim, H. S. Seo, H.-S. Choi *et al.*, *Trichosanthes kirilowii* Ethanol Extract and Cucurbitacin D Inhibit Cell Growth and Induce Apoptosis through Inhibition of STAT3 Activity in Breast Cancer Cells. *Evidence-Based Complementary and Alternative Medicine* **2013**, 1-9 (2013).
133. X.-Q. Zhang, F. C. F. Ip, D.-M. Zhang *et al.*, Triterpenoids with neurotrophic activity from *Ganoderma lucidum*. *Natural Product Research* **25**, 1607-1613 (2011).
134. T. Kawakami, Y. Kawakami and J. Kitaura, Protein Kinase C β (PKC β): Nomal Functions and Dieases. *Journal of Biochemistry* **132**, 677-682 (2002).
135. M. Kashiwagi, M. Ohba, K. Chida and T. Kuroki, Protein Kinase C η (PKC η): Its Involvement in Keratinocyte Differentiation. *Journal of Biochemistry* **132**, 853-857 (2002).
136. N. Saito and Y. Shirai, Protein Kinase C γ (PKC γ): Function of Neuron Specific Isotype. *Journal of Biochemistry* **132**, 683-687 (2002).
137. R. Karodi, M. Jadhav, R. Rub and A. Bafna, Evaluation of the wound healing activity of a crude extract of *Rubia cordifolia* L. (Indian madder) in mice. *International Journal of Applied Research in Natural Products* **2**, 12-18 (2009).
138. L. Cao, Y. Zou, J. Zhu, X. Fan and J. Li, Ginsenoside Rg1 attenuates concanavalin A-induced hepatitis in mice through inhibition of cytokine secretion and lymphocyte infiltration. *Molecular and Cellular Biochemistry* **380**, 203-210 (2013).
139. R. X. Zhang, Z. P. Jia, L. Y. Kong *et al.*, Stachyose extract from *Rehmannia glutinosa* Libosch. to lower plasma glucose in normal and diabetic rats by oral administration. *Pharmazie* **59**, 552-556 (2004).
140. W. L. Li, H. C. Zheng, J. Bukuru and N. D. Kimpe, Natural medicines used in the traditional Chinese medical system for therapy of diabetes mellitus. *Journal of Ethnopharmacology* **92**, 1-21 (2004).
141. W. L. W. Hsiao and L. Liu, The Role of Traditional Chinese Herbal Medicines in Cancer Therapy – from TCM Theory to Mechanistic Insights. *Planta Medica* **76**, 1118-1131 (2010).
142. V. I. Polshakov, Dihydrofolate reductase: structural aspects of mechanism of enzyme catalysis and inhibition. *Russian Chemical Bulletin* **50**, 1733-1751 (2001).
143. F.-M. Hsu, S. Zhang and B. P. C. Chen, Role of DNA-dependent protein kinase catalytic subunit in cancer development and treatment. *Translational Cancer Research* **1**, 22-34 (2012).
144. P. E. Wakefield, W. D. James, C. P. Samlaska and M. S. Meltzer, Tumor necrosis factor *Journal of the American Academy of Dermatology* **24**, 675-685 (1991).
145. W.-M. Chu, Tumor necrosis factor. *Cancer Letters* **328**, 222-225 (2013).
146. M. Adetumbi, G. T. Javor and B. H. S. Lau, *Allium sativum* (Garlic) Inhibits Lipid Synthesis by *Candida albicans*. *Antimicborial Agents and Chemotherapy* **30**, 499-501 (1986).

147. H.-D. Ma, Y.-R. Deng, Z. Tian and Z.-X. Lian, Traditional Chinese Medicine and Immune Regulation. *Clinical Review Allergy Immunology* **44**, 229-241 (2013).
148. S.-L. Tan and P. J. Parker, Emerging and diverse roles of protein kinase C in immune cell signalling. *Biochemical Journal* **376**, 545-552 (2003).
149. R. P. Heaney and C. M. Weaver, Newer Perspectives on Calcium Nutrition and Bone Quality. *Journal of the American College of Nutrition* **24**, 574S-581S (2005).
150. M. Bielohuby, M. Matsuura, N. Herbach *et al.*, Short-term exposure to low-carbohydrate, high-fat diets induces low bone mineral density and reduces bone formation in rats. *Journal of Bone and Mineral Research* **25**, 275-284 (2010).
151. R. E. Mebius and G. Kraal, Structure and function of the spleen. *Nature Reviews Immunology* **5**, 606-616 (2005).
152. F. Zhang, N. K. Altorki, Y.-C. Wu *et al.*, Duodenal reflux induces cyclooxygenase-2 in the esophageal mucosa of rats: Evidence for involvement of bile acids. *Gastroenterology* **121**, 1391-1399 (2001).
153. D.-W. Perng, K.-T. Chang, K.-C. Su *et al.*, Exposure of airway epithelium to bile acids associated with gastroesophageal reflux symptoms: a relation to transforming growth factor- β 1 production and fibroblast proliferation. *Chest* **132**, 1548-1556 (2007).
154. C.-H. Guo, P.-J. Liu, S. Hsia, C.-J. Chuang and P.-C. Chen, Role of certain trace minerals in oxidative stress, inflammation, CD4/CD8 lymphocyte ratios and lung function in asthmatic patients. *Annals of Clinical Biochemistry* **48**, 344-351 (2011).
155. W. Morse, *TCM Treatment of Bronchiol Asthma in Clinical Practice*, (http://www.acupuncture.com/newsletters/m_feb08/TCM_and_Asthma.htm), (2008).
156. N. Bishop, G. J. Walker, M. Gleeson, F. A. Wallace and C. R. Hewitt, Human T lymphocyte migration towards the supernatants of human rhinovirus infected airway epithelial cells: influence of exercise and carbohydrate intake. *Exercise Immunology Review* **15**, 42-59 (2009).
157. M. P. Rayman, The importance of selenium to human health. *The Lancet* **356**, 233-241 (2000).
158. M. Orth and S. Bellosa, Cholesterol: Its Regulation and Role in Central Nervous System Disorders. *Cholesterol* **2012**, 1-19 (2012).
159. Y. Lim, M. Levy and T. M. Bray, Dietary Zinc Alters Early Inflammatory Responses during Cutaneous Wound Healing in Weanling CD-1 Mice. *Journal of Nutrition* **134**, 811-816 (2004).
160. Y.-J. Shiao, J.-C. Chen, C.-N. Wang and C.-T. Wang, The mode of action of primary bile salts on human platelets. *Biochimica et Biophysica Acta* **46**, 282-293 (1993).
161. R. K. Rao and G. Samak, Bile duct epithelial tight junctions and barrier function. *Tissue Barriers* **1**, e25718 (2013).
162. D. Chaudhary, R. Sharma and D. Bansal, Implications of Magnesium Deficiency in Type 2 Diabetes: A Review. *Biological Trace Element Research* **134**, 119-129 (2010).
163. J. Guo, H. Chen, J. Song *et al.*, Syndrome Differentiation of Diabetes by the Traditional Chinese Medicine according to Evidence-Based Medicine and Expert Consensus Opinion. *Evidence-Based Complementary and Alternative Medicine* **2014**, 1-7 (2014).
164. J. F. Oram and A. M. Vaughan, ATP-Binding Cassette Cholesterol Transporters and Cardiovascular Disease. *Circulation Research* **99**, 1031-1043 (2006).
165. M. Blanc, W. Y. Hsieh, K. A. Robertson *et al.*, Host Defense against Viral Infection Involves Interferon Mediated Down-Regulation of Sterol Biosynthesis. *PLoS Biology* **9**, e1000598 (2011).
166. J. L. Wylie, G. M. Hatch and G. McClarty, Host cell phospholipids are trafficked to and then modified by Chlamydia trachomatis. *American Society for Microbiology* **179**, 7233-7242 (1997).
167. D. W. Good, T. George and B. A. Watts, Lipopolysaccharide directly alters renal tubule transport through distinct TLR4-dependent pathways in basolateral and apical membranes. *American Journal of Physiology - Renal Physiology* **297**, F866-F874 (2009).
168. A. Cuenda and S. Rousseau, p38 MAP-Kinases pathway regulation, function and role in human diseases. *Biochimica et Biophysica Acta* **1773**, 1358-1375 (2007).
169. Y. Li and P. Agarwal, A Pathway-Based View of Human Diseases and Disease Relationships. *PLoS ONE* **4**, e4346 (2009).

170. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, The Protein Kinase Complement of the Human Genome. *Science* **298**, 1912-1934 (2002).
171. K. L. Pierce, R. T. Premont and R. J. Lefkowitz, Seven-transmembrane receptors. *Nature* **3**, 640-650 (2002).
172. T. M. Ehrman, D. J. Barlow and P. J. Hylands, Phytochemical Informatics of Traditional Chinese Medicine and Therapeutic Relevance. *Journal of Chemical Information and Modeling* **47**, 2316-2334 (2007).
173. P. A. Williams, J. Cosme, V. Sridhar, E. F. Johnson and D. E. McRee, Mammalian Microsomal Cytochrome P450 Monooxygenase: Structural Adaptations for Membrane Binding and Functional Diversity. *Molecular Cell* **5**, 121-131 (2000).
174. C. T. Supuran and A. Scozzafava, Carbonic anhydrase as targets for medicinal chemistry. *Bioorganic & Medicinal Chemistry* **15**, 4336-4350 (2007).
175. J. L. Boyer, Bile Formation and Secretion. *Comprehensive Physiology* **3**, 1035-1078 (2013).
176. J. Chen, *TCM and Infectious Disease*, (<http://www.acupuncturedtoday.com/mpacms/at/article.php?id=32393>), (2011).
177. A. E. Brady and L. E. Limbird, G protein-coupled receptor interacting proteins: Emerging roles in localization and signal transduction. *Cellular Signalling* **14**, 297-309 (2002).
178. W.-J. Bei, J. Guo, H.-Y. Wu and Y. Cao, Lipid-Regulating Effect of Traditional Chinese Medicine: Mechanisms of Actions. *Evidence-Based Complementary and Alternative Medicine* **2012**, 1-10 (2012).
179. H. K. Ghayee and R. J. Auchus, Basic concepts and recent developments in human steroid hormone biosynthesis. *Reviews in Endocrine and Metabolic Disorders* **8**, 289-300 (2007).
180. M. C. Garcia-Alvarez, I. Moussa, P. Njomnang Soh *et al.*, Both plants *Sebastiania chamaelea* from Niger and *Chrozophora senegalensis* from Senegal used in African traditional medicine in malaria treatment share a same active principle. *Journal of Ethnopharmacology* **149**, 676-684 (2013).
181. M.-J. R. Howes and P. J. Houghton, Plants used in Chinese and Indian traditional medicine for improvement of memory and cognitive function. *Pharmacology Biochemistry and Behavior* **75**, 513-527 (2003).
182. C. C. Lee and P. Houghton, Cytotoxicity of plants from Malaysia and Thailand used traditionally to treat cancer. *Journal of Ethnopharmacology* **100**, 237-243 (2005).
183. D. E. Fairbrothers, T. J. Mabry, R. L. Scogin and B. L. Turner, The Bases of Angiosperm Phylogeny: Chemotaxonomy. *Annals of the Missouri Botanical Garden* **62**, 765-800 (1975).
184. N. Rønsted, V. Savolainen, P. Mølgaard and A. K. Jäger, Phylogenetic selection of Narcissus species for drug discovery. *Biochemical Systematics and Ecology* **36**, 417-422 (2008).
185. M. Wink and G. I. A. Mohamed, Evolution of chemical defense traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the rbcL gene. *Biochemical Systematics and Ecology* **31**, 897-917 (2003).
186. E. Pichersky and E. Lewinsohn, Convergent Evolution in Plant Specialized Metabolism. *Annual Review of Plant Biology* **62**, 549-566 (2011).
187. D.-X. Kong, X.-J. Li and H.-Y. Zhang, Convergent Evolution of Medicines. *ChemMedChem* **3**, 1169-1171 (2008).
188. L. G. Ranilla, Y.-I. Kwon, E. Apostolidis and K. Shetty, Phenolic compounds, antioxidant activity and *in vitro* inhibitory potential against key enzymes relevant for hyperglycemia and hypertension of commonly used medicinal plants, herbs and spices in Latin America. *Bioresource Technology* **101**, 4676-4689 (2010).
189. F. Lopez-Vallejo, M. A. Giulianotti, R. A. Houghten and J. L. Medina-Franco, Expanding the medicinally relevant chemical space with compound libraries. *Drug Discovery Today* **17**, 718-726 (2012).
190. A. B. Yongye, J. Waddell and J. L. Medina-Franco, Molecular Scaffold Analysis of Natural Products Databases in the Public Domain. *Chemical Biology & Drug Design* **80**, 717-724 (2012).
191. A. Sudhakar, History of Cancer, Ancient and Modern Treatment Methods. *Journal of Cancer Science & Therapy* **1**, 1-4 (2010).
192. Y. Ling, Traditional Chinese medicine in the treatment of symptoms in patients with advanced cancer. *Annals of Palliative Medicine* **2**, 141-152 (2013).

193. S. Jain, V. Gill, N. Vasudeva and N. Singla, Ayurvedic medicines in treatment of cancer. *Journal of Chinese Integrative Medicine* **7**, 1096-1099 (2009).
194. P. Balachandran and R. Govindarajan, Cancer—an ayurvedic perspective. *Pharmacological Research* **51**, 19-30 (2005).
195. P. R. Manohar, Descriptions and Claasifications of Cancer in the Classical Ayurvedic Texts. *Indian Journal of History of Science* **20**, 187-195 (2015).
196. J. A. Jamal, Malay traditional medicine. *Tech Monitor*, 37-49 (2006).
197. A. S. Ahmad, *Warisan perubatan Melayu*. (Dewan Bahasa dan Pustaka, Kementerian Pelajaran Malaysia, 1982).
198. P. Garodia, H. Ichikawa, N. Malani, G. Sethi and B. B. Aggarwal, From Ancient Medicine to Modern Medicine: Ayurvedic Concepts of Health and Their Role in Inflammation and Cancer. *Journal of the Society for Integrative Oncology* **5**, 1-16 (2009).
199. J. Duke, *Dr Duke's Phytochemical and Ethnobotanical Databases*, (<https://phytochem.nal.usda.gov/phytochem/search/list>), (1995).
200. D. S. Wishart, T. Jewison, A. C. Guo *et al.*, HMDB 3.0-The Human Metabolome Database 2013. *Nucleic Acids Research* **41**, D801-D807 (2013).
201. H. A. Wahab, *Natural Product Discovery System (NADI)*, (<http://www.nadi-discovery.com/>), (2007).
202. E. W. Sayers, T. Barrett, D. A. Benson *et al.*, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **37**, D5-D15 (2009).
203. phyloT, *A phylogenetic tree generator, based on NCBI taxonomy*, (<http://phylot.biobyte.de/index.html>), (2015).
204. I. Letunic and P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**, W242–W245 (2016).
205. *RDKit: Cheminformatics and Machine Learning Software*, (<http://www.rdkit.org>), (2013).
206. M. R. Berthold, N. Cebron, F. Dill *et al.*, KNIME – The Konstanz Information Miner, Version 2.0 and Beyond. *SIGKDD Explorations Newsletter* **11**, 26-31 (2009).
207. T. Sander, J. Freyss, M. von Korff and C. Rufener, DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *Journal of Chemical Information and Modeling* **55**, 460-473 (2015).
208. G. W. Bemis and M. A. Murcko, The Properties of Known Drugs. 1. Molecular Frameworks. *Journal of Medicinal Chemistry* **39**, 2887-2893 (1996).
209. J. L. Medina-Franco, K. Martinez-Mayorga, A. Bender and T. Scior, Scaffold Diversity Analysis of Compound Data Sets Using an Entropy-Based Measure. *QSAR & Combinatorial Science* **28**, 1551-1560 (2009).
210. ChemAxon Standardizer (6.0.2), (<http://www.chemaxon.com>), (2013).
211. L. H. Mervin, A. M. Afzal, G. Drakakis *et al.*, Target prediction utilising negative bioactivity data covering large chemical space. *Journal of Cheminformatics* **7**, 1-16 (2015).
212. Y. Wang, J. Xiao, T. O. Suzek *et al.*, PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research* **37**, W623–W633 (2009).
213. L. H. Mervin, K. C. Bulusu, L. Kalash *et al.*, Orthologue chemical space and its influence on target prediction. *Bioinformatics*, btx525 (2017).
214. J. C. Platt, Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in Large Margin Classifiers* **10**, 61-74 (1999).
215. *The CTTV Target Validation Platform*, (<http://www.targetvalidation.org>), (2015).
216. C. H. Saslis-Lagoudakis, B. B. Klitgaard, F. Forest *et al.*, The Use of Phylogeny to Interpret Cross-Cultural Patterns in Plant Use and Guide Medicinal Plant Discovery: An Example from *Pterocarpus* (Leguminosae). *PLoS ONE* **6**, e22275 (2011).
217. S. López, J. Bastida, F. Viladomat and C. Codina, Acetylcholinesterase inhibitory activity of some Amaryllidaceae alkaloids and Narcissus extracts. *Life Sciences* **71**, 2521-2529 (2002).
218. M. Wink, Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3-19 (2003).
219. A. Wadood, M. Ghufuran, S. B. Jamal *et al.*, Phytochemical Analysis of Medicinal Plants Occurring in Local Area of Mardan. *Biochemistry & Analytical Biochemistry* **2**, 1-4 (2013).
220. S. Jasial, Y. Hu, M. Vogt and J. Bajorath, Activity-relevant similarity values for fingerprints and implications for similarity searching. *FI000Research* **5**(Chem Inf Sci), 1-14 (2016).

221. F. M. Huennekens, The methotrexate story: A paradigm for development of cancer chemotherapeutic agents. *Advances in Enzyme Regulation* **34**, 397-419 (1994).
222. R. Leuchtenberger, C. Leuchtenberger, D. Laszlo and R. Lewisohn, The Influence of "Folic Acid" on Spontaneous Breast Cancers in Mice. *Science* **101**, 46 (1945).
223. F. Arcamone, G. Cassinelli, G. Fantini *et al.*, Adriamycin, 14-hydroxydaimomycin, a new antitumor antibiotic from *S. Peucetius* var. *caesius*. *Biotechnology and Bioengineering* **11**, 1101-1110 (1969).
224. A. Čihák, Biological Effects of 5-Azacytidine in Eukaryotes. *Oncology* **30**, 405-422 (1974).
225. A. Peer, M. Gottfried, V. Sinibaldi *et al.*, Comparison of Abiraterone Acetate Versus Ketoconazole in Patients with Metastatic Castration Resistant Prostate Cancer Refractory to Docetaxel. *The Prostate* **74**, 433-440 (2014).
226. T. A. Yap, C. P. Carden, G. Attard and J. S. de Bono, Targeting CYP17: established and novel approaches in prostate cancer. *Current Opinion in Pharmacology* **8**, 449-457 (2008).
227. L. Yin and Q. Hu, CYP17 inhibitors-abiraterone, C17,20-lyase inhibitors and multi-targeting agents. *Nature Reviews Urology* **11**, 32-42 (2014).
228. Y. H. Ju, L. M. Clausen, K. F. Allred, A. L. Almada and W. G. Helferich, β -Sitosterol, β -Sitosterol Glucoside, and a Mixture of β -Sitosterol and β -Sitosterol Glucoside Modulate the Growth of Estrogen-Responsive Breast Cancer Cells *In Vitro* and in Ovariectomized Athymic Mice. *Journal of Nutrition* **134**, 1145-1151 (2004).
229. A. M. S. Mayer, K. B. Glaser, C. Cuevas *et al.*, The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends in Pharmacological Sciences* **31**, 255-265 (2010).
230. S. M. Ogbourne and P. G. Parsons, The value of nature's natural product library for the discovery of New Chemical Entities: The discovery of ingenol mebutate. *Fitoterapia* **98**, 36-44 (2014).
231. R. R. Sidambaram, M. G. Dinesh and E. T. Jayalaksmi, An *in vitro* study of cytotoxic activity of *Euphorbia hirta* on Hep-2 cell of human epithelioma of larynx. *International Journal of Pharmacy and Pharmaceutical Sciences* **3**, 101-103 (2011).
232. M. A. Jordan and L. Wilson, Microtubules as a target for anticancer drugs. *Nature Reviews Cancer* **4**, 253-265 (2004).
233. M. J. Siddiqui, Z. Ismail, A. F. A. Aisha and A. M. S. A. Majid, Cytotoxic Activity of *Catharanthus roseus* (Apocynaceae) Crude Extracts and Pure Compounds and Pure Compounds Against Human Colorectal Carcinoma Cell Line. *International Journal of Pharmacology* **6**, 43-47 (2010).
234. L. P. Jordheim, D. Durantel, F. Zoulim and C. Dumontet, Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nature Reviews Drug Discovery* **12**, 447-464 (2013).
235. H. Hussain, A. Al-Harrasi, A. Al-Rawahi *et al.*, A fruitful decade from 20015 to 2014 for anthraquinone patents. *Expert Opinion on Therapeutic Patents* **25**, 1053-1064 (2015).
236. A. Thakur, R. Singla and V. Jaitak, Coumarins as anticancer agents: A review on synthetic strategies, mechanism of action and SAR studies. *European Journal of Medicinal Chemistry* **101**, 476-495 (2015).
237. J.-J. Shie and J.-M. Fang, Phosphonate Congeners of Oseltamivir and Zanamivir as Effective Anti-influenza Drugs: Design, Synthesis and Biological Activity. *Journal of the Chinese Chemical Society* **61**, 127-141 (2014).
238. M. Singh, M. Kaur and O. Silakari, Flavones: An important scaffold for medicinal chemistry. *European Journal of Medicinal Chemistry* **84**, 206-239 (2014).
239. T. A. Woyengo, V. R. Ramprasath and P. J. H. Jones, Anticancer effects of phytosterols. *European Journal of Clinical Nutrition* **63**, 813-820 (2009).
240. E. E. Rufino-Palomares, A. Pérez-Jiménez, F. J. Reyes-Zurita *et al.*, Anti-Cancer and Anti-Angiogenic Properties of Various Natural Pentacyclic Triterpenoids and some of their Chemical Derivatives. *Current Organic Chemistry* **19**, 919-947 (2015).
241. M. M. M. Pinto, M. E. Sousa and M. S. J. Nascimento, Xanthone Derivatives: New Insights in Biological Activities. *Current Medicinal Chemistry* **12**, 2517-2538 (2005).
242. C.-N. Lin, S.-J. Liou, T.-H. Lee, Y.-C. Chuang and S.-J. Won, Xanthone Derivatives as Potential Anti-cancer Drugs. *Journal of Pharmacy and Pharmacology* **48**, 539-544 (1996).
243. C. U. Kim, X. Chen and D. B. Mendel, Neuraminidase inhibitors as anti-influenza virus agents. *Antiviral Chemistry & Chemotherapy* **10**, 141-154 (1999).

244. X. Shang, X. He, X. He *et al.*, The genus *Scutellaria* an ethnopharmacological and phytochemical review. *Journal of Ethnopharmacology* **128**, 279-313 (2010).
245. C. Tsatsanis and D. Spandidos, The role of oncogenic kinases in human cancer (Review). *International Journal of Molecular Medicine* **5**, 583-673 (2000).
246. B. Barneda-Zahonero and M. Parra, Histone deacetylases and cancer. *Molecular Oncology* **6**, 579-589 (2012).
247. S. Ropero and M. Esteller, The role of histone deacetylases (HDACs) in human cancer. *Molecular Oncology* **1**, 19-25 (2007).
248. D. Sarrouilhe, J. Clarhaut, N. Defamie and M. Mesnil, Serotonin and Cancer: What Is the Link? *Current Molecular Medicine* **15**, 62-77 (2015).
249. S. S. Lange, K.-i. Takata and R. D. Wood, DNA polymerases and cancer. *Nature Reviews Cancer* **11**, 96-110 (2011).
250. World Health Organization Technical Report Series, Media Center: Cardiovascular diseases (CVDs), (<http://www.who.int/mediacentre/factsheets/fs317/en/>), (2017).
251. A. Handerson, Coronary heart disease: Overview. *The Lancet* **348**, s1-s4 (1996).
252. C. J. Pepine and W. W. Nichols, The pathophysiology of chronic ischemic heart disease. *Clinical Cardiology* **30**, I-4-I-9 (2007).
253. E. G. Nabel and E. Braunwald, A Tale of Coronary Artery Disease and Myocardial Infarction. *New England Journal of Medicine* **366**, 54-63 (2012).
254. M. Simons and J. A. Ware, Therapeutic angiogenesis in cardiovascular disease. *Nature Reviews Drug Discovery* **2**, 863-872 (2003).
255. L. Deveza, J. Choi and F. Yang, Therapeutic Angiogenesis for Treating Cardiovascular Diseases. *Theranostics* **2**, 801-814 (2012).
256. H. Guo and G. Cheng, Effects of Shexiang Baoxin Pill and Isosorbide Dinitrate on Angina of Coronary Heart Disease: A Meta-Analysis. *Journal of Pharmaceutical and Biomedical Sciences* **16**, 557-563 (2016).
257. P. Jiang, R. Liu, S. Dou *et al.*, Analysis of the constituents in rat plasma after oral administration of Shexiang Baoxin pill by HPLC-ESI-MS/MS. *Biomedical Chromatography* **23**, 1333-1343 (2009).
258. S. Yan, Y. Yang, Y. Wu, R. Liu and W. Zhang, Chemical fingerprinting and quantitative analysis of volatiles in Shexiang Baoxin Pill by gas chromatography with flame ionization and mass spectrometric detection. *Journal of Analytical Chemistry* **64**, 149-155 (2009).
259. S.-K. Yan, W.-D. Zhang, R.-H. Liu and Y.-C. Zhan, Chemical Fingerprinting of *Shexiang Baoxin Pill* and Simultaneous Determination of Its Major Constituents by HPLC with Evaporative Light Scattering Detection and Electrospray Mass Spectrometric Detection. *Chemical and Pharmaceutical Bulletin* **54**, 1058-1062 (2006).
260. L. Xiang, P. Jiang, C. Zhan *et al.*, The serum metabolomic study of intervention effects of the traditional Chinese medicine Shexiang Baoxin Pill and a multi-component medicine polypill in the treatment of myocardial infarction in rats. *Molecular BioSystems* **8**, 2434-2442 (2012).
261. L. Xiang, P. Jiang, S. Wang *et al.*, Metabolomic Strategy for Studying the Intervention and the Synergistic Effects of the Shexiang Baoxin Pill for Treating Myocardial Infarction in Rats. *Evidence-Based Complementary and Alternative Medicine* **2013**, 1-11 (2013).
262. P. Jiang, P. Fu, L. Xiang *et al.*, The effectiveness of borneol on pharmacokinetics changes of four ginsenosides in Shexiang Baoxin Pill *in vivo*. *Biomedical Chromatography* **28**, 419-427 (2014).
263. W.-Y. Jiang, Therapeutic wisdom in traditional Chinese medicine: a perspective from modern science. *Trends in Pharmacological Sciences* **26**, 558-563 (2005).
264. X. Li, G. Qin, Q. Yang, L. Chen and L. Xie, Biomolecular Network-Based Synergistic Drug Combination Discovery. *BioMed Research International* **2016**, 1-11 (2016).
265. M. C. Berenbaum, What is synergy? *Pharmacological Reviews* **41**, 93-141 (1989).
266. S. Loewe, The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* **3**, 285-290 (1953).
267. C. I. Bliss, The toxicity of poisons applied jointly. *Annals of Applied Biology* **26**, 585-615 (1939).
268. K. C. Bulusu, R. Guha, D. J. Mason *et al.*, Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discovery Today* **21**, 225-238 (2016).

269. D. Chen, X. Liu, Y. Yang, H. Yang and P. Lu, Systematic synergy modeling: understanding drug synergy from a systems biology perspective. *BMS Systems Biology* **9**, 1-10 (2015).
270. S. Li, B. Zhang and N. Zhang, Network target for screening synergistic drug combinations with application to traditional Chinese medicine. *BMC Systems Biology* **5**, 1-13 (2011).
271. L. Huang, F. Li, J. Sheng *et al.*, DrugComboRanker: drug combination discovery based on target network analysis. *Bioinformatics* **30**, i228-i236 (2014).
272. E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider and A. Bairoch, UniProtKB/Swiss-Prot: The Manually Annotated Section of the UniProt KnowledgeBase. *Methods in Molecular Biology* **406**, 89-112 (2007).
273. A. P. Davis, C. J. Grondin, K. Lennon-Hopkins *et al.*, The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Research* **43**, D914-D920 (2015).
274. Gene Ontology Consortium: going forward. *Nucleic Acids Research* **43**, D1049-D1056 (2015).
275. G. Wu, X. Feng and L. Stein, A human functional protein interaction network and its application to cancer data analysis. *Genome Biology* **11**, 1-23 (2010).
276. P. Shannon, A. Markiel, O. Ozier *et al.*, Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498-2504 (2003).
277. G. Csardi and T. Nepusz, The igraph software package for complex network research. *InterJournal Complex Systems*, 1695 (2006).
278. G. Bindea, B. Mlecnik, H. Hackl *et al.*, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093 (2009).
279. L. He, E. Kuleskiy, J. Saarela *et al.*, Methods for High-Throughput Drug Combination Screening and Synergy Scoring. *BioRxiv*, 1-14 (2016).
280. P. Libby, A. H. Lichtman and G. K. Hansson, Immune Effector Mechanisms Implicated in Atherosclerosis: From Mice to Humans. *Immunity* **38**, 1092-1104 (2013).
281. X. Meng, J. Yang, M. Dong *et al.*, Regulatory T cells in cardiovascular diseases. *Nature Reviews Cardiology* **13**, 167-179 (2016).
282. F. Mor, F. J. Quintana and I. R. Cohen, Angiogenesis-Inflammation Cross-Talk: Vascular Endothelial Growth Factor Is Secreted by Activated T Cells and Induces Th1 Polarization. *Journal of Immunology* **172**, 4618-4623 (2004).
283. B. A. Rose, T. Force and Y. Wang, Mitogen-Activated Protein Kinase Signaling in the Heart: Angels Versus Demons in a Heart-Breaking Tale. *Physiological Reviews* **90**, 1507-1546 (2010).
284. S. Dhingra, A. K. Sharma, R. C. Arora, J. Slezak and P. K. Singal, IL-10 attenuates TNF- α -induced NF κ B pathway activation and cardiomyocyte apoptosis. *Cardiovascular Research* **82**, 59-66 (2009).
285. M. Newman, The Structure and Function of Complex Networks. *SIAM Review* **45**, 167-256 (2003).
286. K.-i. Kim, H.-J. Cho, J.-Y. Hahn *et al.*, β -Catenin Overexpression Augments Angiogenesis and Skeletal Muscle Regeneration Through Dual Mechanism of Vascular Endothelial Growth Factor-Mediated Endothelial Cell Proliferation and Progenitor Cell Mobilization. *Arteriosclerosis, Thrombosis, and Vascular Biology* **26**, 91-98 (2005).
287. J. Jia, T. Ye, P. Cui *et al.*, AP-1 transcription factor mediates VEGF-induced endothelial cell migration and proliferation. *Microvascular research* **105**, 103-108 (2016).
288. G. P. van Nieuw Amerongen, P. Koolwijk, A. Versteilen and V. W. M. van Hinsbergh, Involvement of RhoA/Rho Kinase Signaling in VEGF-Induced Endothelial Cell Migration and Angiogenesis *In Vitro*. *Arteriosclerosis, Thrombosis, and Vascular Biology* **23**, 211-217 (2003).
289. K. Issbrücker, H. H. Marti, S. Hippenstiel *et al.*, p38 MAP Kinase - a molecular switch between VEGF-induced angiogenesis and vascular hyperpermeability. *FASEB Journal* **17**, 262-264 (2003).
290. A. G. Bader, S. Kang and P. K. Vogt, Cancer-specific mutations in PIK3CA are oncogenic *in vivo*. *Proceedings of the National Academy of Sciences* **103**, 1475-1479 (2006).
291. S. P. Tabruyn and A. W. Griffioen, NF- κ B: a new player in angiostatic therapy. *Angiogenesis* **11**, 101-106 (2008).

292. P. R. Somanath, O. V. Razorenova, J. Chen and T. V. Byzova, Akt1 in Endothelial Cell and Angiogenesis. *Cell cycle* **5**, 512-518 (2006).
293. K. H. Wrighton, Cell signalling: EGF signalling - it's all in SHC1's timing. *Nature Reviews Cell Biology* **14**, 463 (2013).
294. N. Makki, K. W. Thiel and F. J. Miller, The Epidermal Growth Factor Receptor and Its Ligands in Cardiovascular Disease. *International Journal of Molecular Sciences* **14**, 20597-20613 (2013).
295. E. R. Block, M. A. Tolino and J. K. Klarlund, Pyk2 Activation Triggers Epidermal Growth Factor Receptor Signaling and Cell Motility after Wounding Sheets of Epithelial Cells. *Journal of Biological Chemistry* **285**, 13372-13379 (2010).
296. L.-H. Chu, C. G. Rivera, A. S. Popel and J. S. Bader, Constructing the angiome: a global angiogenesis protein interaction network. *Physiological Genomics* **44**, 915-924 (2012).
297. R. Montañez, F. Sánchez-Jiménez, A. R. Quesada and M. Á. Medina, Exploring and challenging the network of angiogenesis. *Scientific Reports* **1**, 1-6 (2011).
298. J. Gu, Y. Chen, S. Li and Y. Li, Identification of responsive gene modules by network-based gene clustering and extending: application to inflammation and angiogenesis. *BMC Systems Biology* **4**, 1-18 (2010).
299. A. Camargo and F. Azuaje, Linking Gene Expression and Functional Network Data in Human Heart Failure. *PLoS ONE* **2**, e1347 (2007).
300. Y.-Y. Wang, K.-J. Xu, J. Song and X.-M. Zhao, Exploring drug combinations in genetic interaction network. *BMC Bioinformatics* **13**, 1-7 (2012).
301. J.-r. Zhu, Y.-f. Tao, S. Lou and Z.-m. Wu, Protective effects of ginsenoside Rb3 on oxygen and glucose deprivation-induced ischemic injury in PC12 cells. *Acta Pharmacologica Sinica* **31**, 273-280 (2010).
302. S. Yoshida, K.-i. Aihara, Y. Ikeda *et al.*, Androgen receptor promotes sex-independent angiogenesis in response to ischemia and is required for activation of vascular endothelial cell growth factor receptor signaling. *Circulation* **128**, 60-71 (2013).
303. S. Kousteni, T. Bellido, L. I. Plotkin *et al.*, Nongenotropic, Sex-Nonspecific Signaling through the Estrogen or Androgen Receptors: Dissociation from Transcriptional Activity. *Cell* **104**, 719-730 (2001).
304. B. R. Walker, Glucocorticoids and Cardiovascular Disease. *European Journal of Endocrinology* **157**, 545-559 (2007).
305. S. Mahmoodzadeh, J. Leber, X. Zhang *et al.*, Cardiomyocyte-specific Estrogen Receptor Alpha Increases Angiogenesis, Lymphangiogenesis and Reduces Fibrosis in the Female Mouse Heart Post-Myocardial Infarction. *Journal of cell science & therapy* **5**, 1-10 (2014).
306. A. A. Lanahan, D. Lech, A. Dubrac *et al.*, PTP1b Is a Physiologic Regulator of Vascular Endothelial Growth Factor Signaling in Endothelial Cells. *Circulation* **130**, 902-909 (2014).
307. V. Sørensen, Y. Zhen, M. Zakrzewska *et al.*, Phosphorylation of Fibroblast Growth Factor (FGF) Receptor 1 at Ser777 by p38 Mitogen-Activated Protein Kinase Regulates Translocation of Exogenous FGF1 to the Cytosol and Nucleus. *Molecular and Cellular Biology* **28**, 4129-4141 (2008).
308. S. Choi, Epidermis proliferative effect of the Panax ginseng Ginsenoside Rb2. *Archives of Pharmacol Research* **25**, 71-76 (2002).
309. L. Lai, B. L. Bohnsack, K. Niederreither and K. K. Hirschi, Retinoic acid regulates endothelial cell proliferation during vasculogenesis. *Development* **130**, 6465-6474 (2003).
310. B. L. Krock, N. Skuli and M. C. Simon, Hypoxia-Induced Angiogenesis: Good and Evil. *Genes & Cancer* **2**, 1117-1133 (2011).
311. L. Kuang, J. Feng, G. He and T. Jing, Knockdown of Nrf2 Inhibits the Angiogenesis of Rat Cardiac Micro-vascular Endothelial Cells under Hypoxic Conditions. *International Journal of Biological Sciences* **9**, 656-665 (2013).
312. M. Potente, L. Ghaeni, D. Baldessari *et al.*, SIRT1 controls endothelial angiogenic functions during vascular growth. *Genes & Development* **21**, 2644-2658 (2007).
313. J. Cai, S. Ahmad, W. G. Jiang *et al.*, Activation of Vascular Endothelial Growth Factor Receptor-1 Sustains Angiogenesis and Bcl-2 Expression Via the Phosphatidylinositol 3-Kinase Pathway in Endothelial Cells. *Diabetes* **52**, 2959-2968 (2003).
314. J. Jia, F. Zhu, X. Ma *et al.*, Mechanisms of drug combinations: interaction and network perspectives. *Nature Reviews Drug Discovery* **8**, 111-129 (2009).

315. F. He, R. Guo, S.-L. Wu, M. Sun and M. Li, Protective Effects of Ginsenoside Rb1 on Human Umbilical Vein Endothelial Cells *In Vitro*. *Journal of Cardiovascular Pharmacology* **50**, 314-320 (2007).
316. G. I. Scott, P. B. Colligan, B. H. Ren and J. Ren, Ginsenosides Rb(1) and Re decrease cardiac contraction in adult rat ventricular myocytes: role of nitric oxide. *British Journal of Pharmacology* **134**, 1159-1165 (2001).
317. A. I. Markowska, K. C. Jefferies and N. Panjwani, Galectin-3 Protein Modulates Cell Surface Expression and Activation of Vascular Endothelial Growth Factor Receptor 2 in Human Endothelial Cells. *Journal of Biological Chemistry* **286**, 29913-29921 (2011).
318. A. Zetser, Y. Bashenko, E. Edovitsky *et al.*, Heparanase Induces Vascular Endothelial Growth Factor Expression: Correlation with p38 Phosphorylation Levels and Src Activation. *Cancer Research* **66**, 1455-1463 (2006).
319. D. Diez, A. M. Wheelock, S. Goto *et al.*, The use of network analyses for elucidating mechanisms in cardiovascular disease. *Molecular BioSystems* **6**, 289-304 (2010).
320. L. W. T. Cheung, K. W. Leung, C. K. C. Wong, R. N. S. Wong and A. S. T. Wong, Ginsenoside-Rg1 induces angiogenesis via non-genomic crosstalk of glucocorticoid receptor and fibroblast growth factor receptor-1. *Cardiovascular Research* **89**, 419-425 (2011).
321. L. d. S. Monteiro, K. X. Bastos, J. Barbosa-Filho *et al.*, Medicinal Plants and Other Living Organisms with Antitumor Potential against Lung Cancer. *Evidence-Based Complementary and Alternative Medicine* **2014**, 1-15 (2014).
322. A. Khodorova, B. Navarro, L. S. Jouaville *et al.*, Endothelin-B receptor activation triggers an endogenous analgesic cascade at sites of peripheral injury. *Nature Medicine* **9**, 1055-1061 (2003).
323. Z.-G. Yang, H.-X. Sun and Y.-P. Ye, Ginsenoside Rd from *Panax notoginseng* Is Cytotoxic towards HeLa Cancer Cells and Induces Apoptosis. *Chemistry & Biodiversity* **3**, 187-197 (2006).
324. K. W. Leung, L. W. T. Cheung, Y. L. Pon *et al.*, Ginsenoside Rb1 inhibits tube-like structure formation of endothelial cells by regulating pigment epithelium-derived factor through the oestrogen β receptor. *British Journal of Pharmacology* **152**, 207-215 (2007).
325. Y. J. Lee, Y. R. Jin, W. C. Lim *et al.*, Ginsenoside Rc and Re stimulate c-Fos expression in MCF-7 human breast carcinoma cells. *Archives of Pharmacal Research* **26**, 53-57 (2003).
326. Y. Nakaya, K. Mawatari, A. Takahashi *et al.*, The phytoestrogen ginsenoside Re activates potassium channels of vascular smooth muscle cells through PI3K/Akt and nitric oxide pathways. *Journal of Medical Investigation* **54**, 381-384 (2007).
327. T. T. Hien, N. D. Kim, Y. R. Pokharel *et al.*, Ginsenoside Rg3 increases nitric oxide production via increases in phosphorylation and expression of endothelial nitric oxide synthase: Essential roles of estrogen receptor-dependent PI3-kinase and AMP-activated protein kinase. *Toxicology and Applied Pharmacology* **246**, 171-183 (2010).
328. K. W. Leung, F. P. Leung, Y. Huang, N. K. Mak and R. N. S. Wong, Non-genomic effects of ginsenoside-Re in endothelial cells via glucocorticoid receptor. *FEBS Letters* **581**, 2423-2428 (2007).
329. Y. Lee, E. Chung, K. Youl Lee *et al.*, Ginsenoside-Rg1, one of the major active molecules from *Panax ginseng*, is a functional ligand of glucocorticoid receptor. *Molecular and Cellular Endocrinology* **133**, 135-140 (1997).

APPENDICES

Chapter 2:

Table A2.1: Top three enriched targets across 14 clusters

Cluster	TCM Therapeutic Action class	TCM Therapeutic Action subclass	Top three enriched targets	Target function reported by literatures
I	Purgative medicinal	Offensive purgative	Protein kinase C beta type	immunomodulation
			Protein kinase C delta type	immunomodulation
			Sodium/glucose cotransporter 2	glucose homeostasis
II	Purgative medicinal	Laxative medicinal	DNA topoisomerase 1	cancer
			Oxysterols receptor LXR-alpha	lipid homeostasis
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
	Hemostatic medicinal	Astringent hemostatic	DNA topoisomerase 1	cancer
			Telomerase reverse transcriptase	cancer
			Sodium/glucose cotransporter 1	glucose homeostasis
III	Anti-malarial medicinal	Anti-malarial medicinal	Telomerase reverse transcriptase	cancer
			DNA topoisomerase 1	cancer
			Wee1-like protein kinase	cancer
	Astringent	Secure essence, reduce urination, and check vaginal discharge	Oxysterols receptor LXR-alpha	lipid homeostasis
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
			DNA topoisomerase 1	cancer
	Dampness resolving	Water draining and anti-icteric	DNA topoisomerase 1	cancer
			Sodium/glucose cotransporter 2	glucose homeostasis
			Sodium/glucose cotransporter 1	glucose homeostasis
	Tonifying and replenishing	Yang tonifying	Sodium/glucose cotransporter 2	glucose homeostasis
			DNA topoisomerase 1	cancer
			Sodium/glucose cotransporter 1	glucose homeostasis
	Heat clearing medicinal	Heat clearing and detoxicating	DNA topoisomerase 1	cancer

IV	Tonifying and replenishing	Yin tonifying	Sodium/glucose cotransporter 2	glucose homeostasis
			Sodium/glucose cotransporter 1	glucose homeostasis
			Sodium/glucose cotransporter 2	glucose homeostasis
	Exterior releasing	Wind heat dispersing	DNA topoisomerase 1	cancer
			Oxysterols receptor LXR-alpha	lipid homeostasis
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
	Wind-dampness dispelling	Heat clearing	DNA topoisomerase 1	cancer
			Sodium/glucose cotransporter 2	glucose homeostasis
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
	Heat clearing medicinal	Heat clearing and fire purging	DNA topoisomerase 1	cancer
			Sodium/glucose cotransporter 1	glucose homeostasis
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
	Dampness resolving	Water draining and strangury resolving	Oxysterols receptor LXR-alpha	lipid homeostasis
			DNA topoisomerase 1	cancer
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
	Heat clearing medicinal	Heat clearing and dampness drying	Sodium/glucose cotransporter 2	glucose homeostasis
			DNA topoisomerase 1	cancer
			Protein kinase C beta type	immunomodulation
	Blood activating and stress resolving	Blood activating trauma curing	Leukotriene B4 receptor 1	lipid homeostasis
			DNA topoisomerase 1	cancer
			Estrogen receptor	cancer
	Liver-pacifying and wind extinguishing	Extinguish wind to resolve convulsion	Oxysterols receptor LXR-alpha	lipid homeostasis
			Leukotriene B4 receptor 1	immunomodulation
			Glutamate carboxypeptidase 2	central nervous system
	Blood activating and stress resolving	Blood activating menstruation resolving	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
			Oxysterols receptor LXR-alpha	lipid homeostasis
			DNA topoisomerase 1	cancer

	Cough suppressing and panting-calming	Cough suppressing and panting calming	Tyrosine-protein phosphatase non-receptor type 2 DNA topoisomerase 1 Leukotriene B4 receptor 1	immunomodulation cancer immunomodulation
	Worm expelling medicinal	Worm expelling medicinal	Leukotriene B4 receptor 1 Tyrosine-protein phosphatase non-receptor type 2 Sodium/glucose cotransporter 2	immunomodulation immunomodulation glucose homeostasis
	Astringent	Lung-intestine astringent	Protein kinase C beta type Endothelin B receptor DNA topoisomerase 1	immunomodulation cardio vasodilation cancer
	Digestant medicinal	Digestant medicinal	Oxysterols receptor LXR-alpha Leukotriene B4 receptor 1 DNA topoisomerase 1	lipid homeostasis immunomodulation cancer
V	Cough suppressing and panting-calming	Cold phlegm resolving and warming	DNA topoisomerase 1 Tyrosine-protein phosphatase non-receptor type 2 Telomerase reverse transcriptase	cancer immunomodulation cancer
VI	Emetic medicinal		Integrin alpha-L Poly [ADP-ribose] polymerase 1 Multidrug resistance-associated protein 1	immunomodulation inflammation cytotoxicity
VII	Heat clearing medicinal	Heat clearing and blood cooling	Protein kinase C beta type DNA topoisomerase 1 Sodium/glucose cotransporter 2	immunomodulation cancer glucose homeostasis
	Hemostatic medicinal	Stasis-resolving hemostatic	Tyrosine-protein phosphatase non-receptor type 2 Protein kinase C eta type Protein kinase C gamma type	immunomodulation immunomodulation immunomodulation
	Tonifying and replenishing	Blood tonifying	Tyrosine-protein phosphatase non-receptor type 2 Protein kinase C beta type Protein kinase C eta type	immunomodulation immunomodulation immunomodulation
	Orifice opening	Orifice opening	Steryl-sulfatase Testosterone 17-beta-dehydrogenase 3	immunomodulation reproduction system
VIII				

			Ileal sodium/bile acid cotransporter	bile
IX	Purgative medicinal	Drastic (purgative) water-expelling	Protein kinase C beta type	immunomodulation
			Protein kinase C delta type	immunomodulation
			Protein kinase C eta type	immunomodulation
	Dampness resolving	Water draining and swelling dispersing	Glutamate carboxypeptidase 2	central nervous system
			Protein kinase C beta type	immunomodulation
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
X	Wind-dampness dispelling	Cold dispersing	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
			DNA topoisomerase 1	cancer
			Sodium/glucose cotransporter 1	glucose homeostasis
	Wind-dampness dispelling	Bone(sinew) strengthening	DNA topoisomerase 1	cancer
			Sodium/glucose cotransporter 1	glucose homeostasis
			Steryl-sulfatase	immunomodulation
	Tonifying and replenishing	Qi tonifying	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
			Sodium/glucose cotransporter 2	glucose homeostasis
			Sodium/glucose cotransporter 1	glucose homeostasis
	Cough suppressing and panting-calming	Clearing and Heat phlegm resolving	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
			DNA topoisomerase 1	cancer
			Testosterone 17-beta-dehydrogenase 3	reproductive system
XI	Tranquilizing	Heat nourishing tranquilizing	Peptidyl-prolyl cis-trans isomerase FKBP1A	immunomodulation
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
			Glutamate carboxypeptidase 2	central nervous system
	Astringent	Anhidrotic	Beta-1 adrenergic receptor	cardiac contraction
			5-hydroxytryptamine receptor 2A	central nervous system
			Glutamate [NMDA] receptor subunit epsilon-2	central nervous system
XII	Parasite destroying, dampness eliminating and itchiness relieving		Dihydrofolate reductase	cancer, bacterial infection
			DNA-dependent protein kinase catalytic subunit	cancer
			Tumor necrosis factor	cancer, bacterial infection
XII	Hemostatic medicinal	Meridian warming hemostatic	Peptidyl-prolyl cis-trans isomerase FKBP1A	immunomodulation

XIV	Exterior releasing	Wind cold dispersing	Multidrug resistance protein 1	cytotoxicity
			Steryl-sulfatase	immunomodulation
			Steryl-sulfatase	immunomodulation
			Endothelin B receptor	cardio vasodilation
	Interior warming	Interior warming	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	immunomodulation
			Protein kinase C delta type	immunomodulation
			Multidrug resistance protein 1	cytotoxicity
			Steroid hormone receptor ERR1	relate to estrogen receptor
	Heat clearing medicinal	Deficiency	Sodium/glucose cotransporter 2	glucose homeostasis
			DNA topoisomerase 1	cancer
			Sodium/glucose cotransporter 1	glucose homeostasis
			DNA topoisomerase 1	cancer
	Qi regulating	Qi regulating	Glutamate receptor, ionotropic kainate 1	central nervous system
			Steryl-sulfatase	immunomodulation
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
	Blood activating and stress resolving	Blood activating analgesic	Tyrosine-protein phosphatase non-receptor type 1	immunomodulation
			Prostaglandin E2 receptor EP2 subtype	bone metabolism
			Oxysterols receptor LXR-alpha	lipid homeostasis
	Hemostatic medicinal	Blood cooling hemostatic	Tyrosine-protein phosphatase non-receptor type 2	immunomodulation
			Leukotriene B4 receptor 1	lipid homeostasis
			Oxysterols receptor LXR-alpha	lipid homeostasis
	Blood activating and stress resolving	Blood breaking mass eliminating	Leukotriene B4 receptor 1	immunomodulation
			Tyrosine-protein phosphatase non-receptor type 2	immunomodulation

Table A2.2: Top three enriched pathways across 14 clusters

Cluster	TCM Therapeutic Action class	TCM Therapeutic Action subclass	Top three enriched pathways	Pathway motif according to KEGG
I	Purgative medicinal	Offensive purgative	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04530 Tight junction	Cellular commiunity
II	Purgative medicinal	Laxative medicinal	hsa04978 Mineral absorption	Digestive system
			hsa00040 Pentose and glucuronate interconversions	Carbohydrate metabolism
			hsa00561 Glycerolipid metabolism	Lipid metabolism
	Hemostatic medicinal	Astringent hemostatic	hsa04978 Mineral absorption	Digestive system
			hsa00040 Pentose and glucuronate interconversions	Carbohydrate metabolism
			hsa00561 Glycerolipid metabolism	Lipid metabolism
	Liver-pacifying and wind extinguishing	Liver yang calming	hsa04978 Mineral absorption	Digestive system
			hsa00040 Pentose and glucuronate interconversions	Carbohydrate metabolism
			hsa00561 Glycerolipid metabolism	Lipid metabolism
III	Anti-malarial medicinal	Anti-malarial medicinal	hsa04978 Mineral absorption	Digestive system
			hsa00040 Pentose and glucuronate interconversions	Carbohydrate metabolism
			hsa00561 Glycerolipid metabolism	Lipid metabolism
	Astringent	Secure essence, reduce urination, and check vaginal discharge	hsa04978 Mineral absorption	Digestive system
			hsa00040 Pentose and glucuronate interconversions	Carbohydrate metabolism
			hsa00561 Glycerolipid metabolism	Lipid metabolism
	Dampness resolving	Water draining and anti-icteric	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa00040 Pentose and glucuronate interconversions	Carbohydrate metabolism
	Tonifying and replenishing	Yang tonifying	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa00040 Pentose and glucuronate interconversions	Carbohydrate metabolism
	Heat clearing medicinal	Heat clearing and detoxicating	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
	Tonifying and replenishing	Yin tonifying	hsa04978 Mineral absorption	Digestive system

IV	Exterior releasing	Wind heat dispersing	hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
			hsa04978 Mineral absorption	Digestive system
	Wind-dampness dispelling	Heat clearing	hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
			hsa04978 Mineral absorption	Digestive system
	Heat clearing medicinal	Heat clearing and fire purging	hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
			hsa04978 Mineral absorption	Digestive system
	Dampness resolving	Water draining and strangury resolving	hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
			hsa04978 Mineral absorption	Digestive system
	Heat clearing medicinal	Heat clearing and dampness drying	hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa04978 Mineral absorption	Digestive system
	Blood activating and stress resolving	Blood activating trauma curing	hsa04976 Bile secretion	Digestive system
			hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
	Liver-pacifying and wind extinguishing	Extinguish wind to resolve convulsion	hsa04976 Bile secretion	Digestive system
			hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa04966 Collecting duct acid secretion	Excretory system
	Blood activating and stress resolving	Blood activating menstruation resolving	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
	Cough suppressing and panting-calming	Cough suppressing and panting calming	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system

	Worm expelling medicinal	Worm expelling medicinal	hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
V	Astringent	Lung-intestine astringent	hsa04976 Bile secretion	Digestive system
			hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa04530 Tight junction	Cellular commiunity
	Digestant medicinal	Digestant medicinal	hsa04976 Bile secretion	Digestive system
			hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa00140 Steroid hormone biosynthesis	Lipid metabolism
VI	Cough suppressing and panting-calming	Cold phlegm resolving and warming	hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa00100 Steroid biosynthesis	Lipid metabolism
			hsa04520 Adherens junction	Cellular commiunity
			hsa03410 Base excision repair	Replication and repair
VII	Heat clearing medicinal	Heat clearing and blood cooling	hsa00140 Steroid hormone biosynthesis	Lipid metabolism
			hsa04976 Bile secretion	Digestive system
			hsa04978 Mineral absorption	Digestive system
	Hemostatic medicinal	Stasis-resolving hemostatic	hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa02010 ABC transporters	Membrane transport
			hsa04976 Bile secretion	Digestive system
	Tonifying and replenishing	Blood tonifying	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
VIII	Orifice opening	Orifice opening	hsa04530 Tight junction	Cellular commiunity
			hsa00140 Steroid hormone biosynthesis	Lipid metabolism
			hsa04976 Bile secretion	Digestive system
IX	Purgative medicinal	Drastic (purgative) water-expelling	hsa00900 Terpenoid backbone biosynthesis	Metabolism of terpenoids and polyketides
			hsa04978 Mineral absorption	Digestive system

			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04530 Tight junction	Cellular commiunity
	Dampness resolving	Water draining and swelling dispersing	hsa00900 Terpenoid backbone biosynthesis	Metabolism of terpenoids and polyketides
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04530 Tight junction	Cellular commiunity
	Wind-dampness dispelling	Cold dispersing	hsa04530 Tight junction	Cellular commiunity
			hsa05143 African trypanosomiasis	Infectious diseases
			hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
	Wind-dampness dispelling	Bone(sinew) strengthening	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
	Tonifying and replenishing	Qi tonifying	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
X	Cough suppressing and panting-calming	Clearing and Heat phlegm resolving	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa04976 Bile secretion	Digestive system
	Tranquilizing	Heat nourishing tranquilizing	hsa04978 Mineral absorption	Digestive system
			hsa04973 Carbohydrate digestion and absorption	Digestive system
			hsa00900 Terpenoid backbone biosynthesis	Metabolism of terpenoids and polyketides
XI	Astringent	Anhidrotic	hsa04540 Gap junction	Cellular commiunity
			hsa04970 Salivary secretion	Digestive system
			hsa04020 Calcium signaling pathway	Signal transduction
XII	Parasite destroying, dampness eliminating and itchiness relieving		hsa00100 Steroid biosynthesis	Lipid metabolism
			hsa00564 Glycerophospholipid metabolism	Lipid metabolism
			hsa04966 Collecting duct acid secretion	Excretory system
XII	Hemostatic medicinal	Meridian warming hemostatic	hsa00100 Steroid biosynthesis	Lipid metabolism
			hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa05143 African trypanosomiasis	Infectious diseases

XIV	Exterior releasing	Wind cold dispersing	hsa00100 Steroid biosynthesis	Lipid metabolism
			hsa00564 Glycerophospholipid metabolism	Lipid metabolism
			hsa00140 Steroid hormone biosynthesis	Lipid metabolism
	Interior warming	Interior warming	hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa05143 African trypanosomiasis	Infectious diseases
			hsa05110 Vibrio cholerae infection	Infectious diseases
	Heat clearing medicinal	Deficiency	hsa04978 Mineral absorption	Digestive system
			hsa00040 Pentose and glucuronate interconversions	Carbohydrate metabolism
			hsa00561 Glycerolipid metabolism	Lipid metabolism
	Qi regulating	Qi regulating	hsa00100 Steroid biosynthesis	Lipid metabolism
			hsa00140 Steroid hormone biosynthesis	Lipid metabolism
			hsa04612 Antigen processing and presentation	Immune system
XIV	Blood activating and stress resolving	Blood activating analgesic	hsa00100 Steroid biosynthesis	Lipid metabolism
			hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa00140 Steroid hormone biosynthesis	Lipid metabolism
	Hemostatic medicinal	Blood cooling hemostatic	hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa00140 Steroid hormone biosynthesis	Lipid metabolism
			hsa04520 Adherens junction	Cellular community
	Blood activating and stress resolving	Blood breaking mass eliminating	hsa04961 Endocrine and other factor-regulated calcium reabsorption	Excretory system
			hsa00140 Steroid hormone biosynthesis	Lipid metabolism
			hsa05143 African trypanosomiasis	Infectious diseases

Chapter 3:

Table A3.1: The list of 28 anti-cancer compounds that shows to be structurally similar to TM compounds.

	Anti-cancer Drug	DrugBank ID	TM Compound	Plant	Source of TM	Tc value
1	Abiraterone	DB05812	Brassicasterol	<i>Houttuynia cordata</i>	TCM	0.339
				<i>Moringa oleifera</i>	Ayurveda, MalayTM	
1	Abiraterone	DB05812	Beta-Sitosterol	<i>Abrus precatorius</i>	Ayurveda, MalayTM	0.331
				<i>Alpinia officinarum</i>	TCM	
				<i>Angelica sinensis</i>	TCM	
				<i>Artemisia argyi</i>	TCM	
				<i>Astragalus complanatus</i>	TCM	
				<i>Azadirachta indica</i>	Ayurveda	
				<i>Brucea javanica</i>	TCM	
				<i>Capparis spinosa</i>	Ayurveda	
				<i>Carthamus tinctorius</i>	TCM	
				<i>Centipeda minima</i>	TCM	
				<i>Chrysanthemum indicum</i>	TCM	
				<i>Cornus officinalis</i>	TCM	
				<i>Cuscuta chinensis</i>	TCM	
				<i>Drynaria fortunei</i>	TCM	
				<i>Foeniculum vulgare</i>	TCM	
				<i>Forsythia suspensa</i>	TCM	
				<i>Gynostemma pentaphyllum</i>	TCM	
				<i>Imperata cylindrica</i>	TCM	
				<i>Inula britannica</i>	TCM	
				<i>Isatis tinctoria</i>	TCM	
				<i>Lycopus lucidus</i>	TCM	
				<i>Ocimum tenuiflorum</i>	Ayurveda	
				<i>Oldenlandia diffusa</i>	TCM	
				<i>Paeonia suffruticosa</i>	TCM	
				<i>Persicaria orientalis</i>	TCM	
				<i>Polygonum aviculare</i>	TCM	
				<i>Polygonum cuspidatum</i>	TCM	
				<i>Portulaca oleracea</i>	TCM	
				<i>Prunella vulgaris</i>	TCM, MalayTM	
				<i>Pulsatilla chinensis</i>	TCM	
				<i>Raphanus sativus</i>	Ayurveda	
				<i>Ricinus communis</i>	Ayurveda	
				<i>Scutellaria baicalensis</i>	TCM	
				<i>Scutellaria barbata</i>	TCM	
				<i>Sophora japonica</i>	TCM	
				<i>Spatholobus suberectus</i>	TCM	
				<i>Terminalia arjuna</i>	Ayurveda	
				<i>Vitis vinifera</i>	Ayurveda	
				<i>Zanthoxylum nitidum</i>	TCM	
				<i>Zingiber officinale</i>	TCM, Ayurveda	

24-Methylcholesta-5,24-Dien-3Beta-Ol	<i>Withania somnifera</i>	Ayurveda	0.331
Cholesterol	<i>Allium sativum</i>	Ayurveda, MalayTM	0.328
	<i>Vitis vinifera</i>	Ayurveda	
Campesterol	<i>Abrus precatorius</i>	Ayurveda, MalayTM	0.325
	<i>Curcuma longa</i>	TCM, Ayurveda, MalayTM	
	<i>Curcuma longa</i>	TCM	
	<i>Cuscuta chinensis</i>	TCM	
	<i>Senna tora</i>	TCM, MalayTM	
	<i>Spatholobus suberectus</i>	TCM	
	<i>Vitis vinifera</i>	Ayurveda	
(3S,8S,9S,10R,13R,14S,17R)-17-[(1R,4R)-1,4-Dimethylhexyl]-10,13-Dimethyl-2,3,4,7,8,9,11,12,14,15,16,17-Dodecahydro-1H-Cyclopenta[A]Pheanthren-3-Ol	<i>Eclipta prostrata</i>	TCM	0.323
Daucosterol	<i>Spatholobus suberectus</i>	TCM	0.321
Fucosterol	<i>Cymbopogon citratus</i>	Ayurveda, MalayTM	0.320
	<i>Gynostemma pentaphyllum</i>	TCM	
	<i>Cuscuta chinensis</i>	TCM	
	<i>Withania somnifera</i>	Ayurveda	
(3S,8S,9S,10R,13R,14S,17R)-17-[(E,1R,4R)-1,4-Dimethylhex-2-Enyl]-10,13-Dimethyl-2,3,4,7,8,9,11,12,14,15,16,17-Dodecahydro-1H-Cyclopenta[A]Pheanthren-3-Ol	<i>Eclipta prostrata</i>	TCM	0.320
Stigmasterol	<i>Abrus precatorius</i>	Ayurveda, MalayTM	0.315
	<i>Angelica sinensis</i>	TCM	
	<i>Artemisia argyi</i>	TCM	
	<i>Carthamus tinctorius</i>	TCM	
	<i>Centipeda minima</i>	TCM	
	<i>Cornus officinalis</i>	TCM	
	<i>Curcuma longa</i>	TCM, Ayurveda, MalayTM	
	<i>Drynaria fortunei</i>	TCM	
	<i>Eurycoma longifolia</i>	MalayTM	
	<i>Foeniculum vulgare</i>	TCM	
	<i>Imperata cylindrica</i>	TCM	

				<i>Isatis tinctoria</i>	TCM	
				<i>Oldenlandia diffusa</i>	TCM	
				<i>Prunella vulgaris</i>	TCM, MalayTM	
				<i>Pulsatilla chinensis</i>	TCM	
				<i>Ricinus communis</i>	Ayurveda	
				<i>Scutellaria barbata</i>	TCM	
				<i>Senna tora</i>	TCM, MalayTM	
				<i>Spatholobus suberectus</i>	TCM	
				<i>Vitis vinifera</i>	Ayurveda	
			Clerosterol	<i>Moringa oleifera</i>	Ayurveda, MalayTM	0.313
2	Aminolevulinic Acid	DB00855	Glycine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.450
				<i>Raphanus sativus</i>	Ayurveda	
				<i>Vitis vinifera</i>	Ayurveda	
			Succinic Acid	<i>Allium sativum</i>	Ayurveda, MalayTM	0.450
				<i>Angelica sinensis</i>	TCM	
				<i>Carthamus tinctorius</i>	TCM	
				<i>Forsythia suspensa</i>	TCM	
				<i>Isatis tinctoria</i>	TCM	
				<i>Oldenlandia diffusa</i>	TCM	
				<i>Portulaca oleracea</i>	TCM	
				<i>Ricinus communis</i>	Ayurveda	
				<i>Spatholobus suberectus</i>	TCM	
				<i>Vitis vinifera</i>	Ayurveda	
			Gamma-Aminobutyric Acid	<i>Annona squamosa</i>	Ayurveda	0.400
				<i>Isatis tinctoria</i>	TCM	
				<i>Vitis vinifera</i>	Ayurveda	
3	Azacitidine	DB00928	Guanosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.407
				<i>Carthamus tinctorius</i>	TCM	
			Adenosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.385
			Uridine	<i>Isatis tinctoria</i>	TCM	0.361
4	Choline C 11	DB09277	Choline	<i>Allium sativum</i>	Ayurveda, MalayTM	1.000
				<i>Capparis spinosa</i>	Ayurveda	
5	Clofarabine	DB00631	Adenosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.367
			Guanosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.344
				<i>Carthamus tinctorius</i>	TCM	0.344
6	Cytarabine	DB00987	Uridine	<i>Isatis tinctoria</i>	TCM	0.420
			Guanosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.357
				<i>Carthamus tinctorius</i>	TCM	
			Adenosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.333
7	Decitabine	DB01262	Thymidine	<i>Eurycoma longifolia</i>	MalayTM	0.368
8	Epirubicin	DB00445	Glutathione	<i>Allium sativum</i>	Ayurveda, MalayTM	0.545
9	Floxuridine	DB00322	Thymidine	<i>Eurycoma longifolia</i>	MalayTM	0.591
10	Fludarabine	DB01073	Adenosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.324
			Guanosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.318
				<i>Carthamus tinctorius</i>	TCM	
11	Hydroxyurea	DB01005	Urea	<i>Aloe vera</i>	Ayurveda, MalayTM	0.455

12	Idarubicin	DB01177	Glutathione	<i>Allium sativum</i>	Ayurveda, MalayTM	0.411
13	Ingenol Mebutate	DB05013	Ingenol- Triacetate	<i>Euphorbia hirta</i>	Ayurveda	0.456
14	Melphalan	DB01042	Tyrosine	<i>Abrus precatorius</i> <i>Raphanus sativus</i> <i>Vitis vinifera</i>	Ayurveda, MalayTM Ayurveda Ayurveda	0.436
			Phenylalanine	<i>Albizia lebbbeck</i> <i>Carthamus tinctorius</i> <i>Raphanus sativus</i> <i>Vitis vinifera</i>	Ayurveda TCM Ayurveda Ayurveda	0.389
			Levodopa	<i>Spatholobus suberectus</i> <i>Portulaca oleracea</i>	TCM	0.317
15	Mesna	DB09110	Taurine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.474
16	Methotrexate	DB00563	Folacin	<i>Aloe vera</i> <i>Raphanus sativus</i> <i>Vitis vinifera</i>	Ayurveda, MalayTM Ayurveda Ayurveda	0.388
17	Methyltestoster one	DB06710	4-Campesten-3- One Stigmast-4-En- 3-One	<i>Melia azedarach</i> <i>Melia azedarach</i> <i>Scutellaria barbata</i> <i>Isatis tinctoria</i>	Ayurveda Ayurveda TCM TCM	0.367 0.360
18	Nelarabine	DB01280	Guanosine	<i>Allium sativum</i> <i>Carthamus tinctorius</i>	Ayurveda, MalayTM TCM	0.456
			Adenosine	<i>Allium sativum</i>	Ayurveda, MalayTM	0.420
19	Pemetrexed	DB00642	Folacin	<i>Aloe vera</i> <i>Raphanus sativus</i> <i>Vitis vinifera</i>	Ayurveda, MalayTM Ayurveda Ayurveda	0.379
20	Pralatrexate	DB06813	Folacin	<i>Aloe vera</i> <i>Raphanus sativus</i> <i>Vitis vinifera</i>	Ayurveda, MalayTM Ayurveda Ayurveda	0.367 0.367 0.367
21	Sodium phenylbutyrate	DB06819	Hydrocinnamic acid	<i>Aloe vera</i> <i>Vitis vinifera</i>	Ayurveda, MalayTM Ayurveda	0.541
			Benzyl-acetone	<i>Aloe vera</i> <i>Alpinia officinarum</i>	Ayurveda, MalayTM TCM	0.439
			Phenylethanol	<i>Lawsonia inermis</i> <i>Vitis vinifera</i>	MalayTM Ayurveda	0.405
			2-phenylethyl- amine	<i>Vitis vinifera</i>	Ayurveda	0.368
			Benzyl alcohol	<i>Albizia lebbbeck</i> <i>Lawsonia inermis</i> <i>Vitis vinifera</i>	Ayurveda MalayTM Ayurveda	0.333
			Tetra- decylbenzene	<i>Aloe vera</i>	Ayurveda, MalayTM	0.327
			Tridecylbenzene	<i>Aloe vera</i>	Ayurveda, MalayTM	0.327
			Phenethyl isobutyrate	<i>Alpinia officinarum</i>	TCM	0.327
			Isoamylbenzene	<i>Angelica sinensis</i>	TCM	0.326
			Phenylalanine	<i>Albizia lebbbeck</i> <i>Carthamus tinctorius</i>	Ayurveda TCM	0.304

				<i>Raphanus sativus</i>	Ayurveda	
				<i>Vitis vinifera</i>	Ayurveda	
			Benzyl-acetate	<i>Melia azedarach</i>	Ayurveda	0.304
			Hyacinthin	<i>Artemisia argyi</i>	TCM	0.300
				<i>Carthamus tinctorius</i>	TCM	
				<i>Cinnamomum aromaticum</i>	TCM	
				<i>Gynostemma pentaphyllum</i>	TCM	
				<i>Magnolia officinalis</i>	TCM	
				<i>Scutellaria barbata</i>	TCM	
	Uridine		Uridine			
22	Triacetate	DB09144	Uridine	<i>Isatis tinctoria</i>	TCM	0.361
23	Valrubicin	DB00385	Glutathione	<i>Allium sativum</i>	Ayurveda, MalayTM	0.446
24	Vinblastine	DB00570	Leurosidine (Vinrosidine)	<i>Catharanthus roseus</i>	MalayTM	1.000
			Vincalukoblast ine		MalayTM	0.883
			Leurocristine		MalayTM	0.758
			Vinblastine (Vincalucoblas tine)		MalayTM	0.702
			3',4'- Anhydrovinblas tine		MalayTM	0.682
			N- Deformylvincris tine		MalayTM	0.634
			Catharine		MalayTM	0.623
			Leurosinone		MalayTM	0.579
			Pleurosine		MalayTM	0.500
			Vincathicine		MalayTM	0.469
			Vindolicine		MalayTM	0.454
			Vindoline		MalayTM	0.409
			Deacetylvindoli ne		MalayTM	0.300
25	Vincristine	DB00541	Leurocristine	<i>Catharanthus roseus</i>	MalayTM	1.000
			Leurosidine (Vinrosidine)		MalayTM	0.758
			Vincalukoblast ine		MalayTM	0.687
			N- Deformylvincris tine		MalayTM	0.652
			Vinblastine (Vincalucoblas tine)		MalayTM	0.525
			3',4'- Anhydrovinblas tine		MalayTM	0.507
			Catharine		MalayTM	0.491
			Leurosinone		MalayTM	0.438
			Pleurosine		MalayTM	0.384
			Vindoline		MalayTM	0.342
			Vindolicine		MalayTM	0.342
			Vincathicine		MalayTM	0.342

26	Vincristine	DB00541	Leurocristine	<i>Catharanthus roseus</i>	MalayTM	1.000
			Leurosidine (Vinrosidine)		MalayTM	0.758
			Vincaleukoblastine		MalayTM	0.687
			N-Deformylvincristine		MalayTM	0.652
			Vinblastine (Vincaleucoblastine)		MalayTM	0.525
			3',4'-Anhydrovinblastine		MalayTM	0.507
			Catharine		MalayTM	0.491
			Leurosinone		MalayTM	0.438
			Pleurosine		MalayTM	0.384
			Vindoline		MalayTM	0.342
			Vindolicine		MalayTM	0.342
			Vincathicine		MalayTM	0.342
27	Vindesine	DB00309	Leurosidine (Vinrosidine)	<i>Catharanthus roseus</i>	MalayTM	0.773
			Vincaleukoblastine		MalayTM	0.754
			Leurocristine		MalayTM	0.641
			N-Deformylvincristine		MalayTM	0.562
			Vinblastine (Vincaleucoblastine)		MalayTM	0.537
			Leurosine		MalayTM	0.537
			3',4'-Anhydrovinblastine		MalayTM	0.518
			Catharine		MalayTM	0.475
			Leurosinone		MalayTM	0.436
			Pleurosine		MalayTM	0.377
			Vincathicine		MalayTM	0.351
			Deacetylvindoline		MalayTM	0.351
			Catharosine		MalayTM	0.349
			Vindolicine		MalayTM	0.307
28	Vinorelbine	DB00361	3',4'-Anhydrovinblastine	<i>Catharanthus roseus</i>	MalayTM	0.765
			Leurosidine (Vinrosidine)		MalayTM	0.603
			Vinblastine (Vincaleucoblastine)		MalayTM	0.583
			Catharine		MalayTM	0.577
			Leurosinone		MalayTM	0.572
			Vincaleukoblastine		MalayTM	0.532
			Pleurosine		MalayTM	0.474

Vincathicine	MalayTM	0.462
Vindolicine	MalayTM	0.461
Leurocristine	MalayTM	0.440
Vindoline	MalayTM	0.400
N-Deformylvincristine	MalayTM	0.363

Chapter 4:

Table A4.1: The list of enriched GO biological processes based on SBP predicted targets.

GOID	GO Term	Term p-value	GO Levels	% Associated Genes
GO:0000187	activation of MAPK activity	65.0E-15	[7, 8, 9, 10, 11, 12, 13]	13.55
GO:0032872	regulation of stress-activated MAPK cascade	270.0E-18	[5, 6, 7, 8, 9, 10, 11]	12.75
GO:0051403	stress-activated MAPK cascade	28.0E-21	[5, 6, 7, 9, 10]	12.10
GO:0002224	toll-like receptor signaling pathway	2.4E-12	[6, 8, 9, 10, 11]	12.05
GO:0043406	positive regulation of MAP kinase activity	7.1E-15	[6, 7, 8, 9, 10, 11, 12]	11.16
	positive regulation of sequence-specific DNA			
GO:0051091	binding transcription factor activity	130.0E-15	[4, 6, 7, 8, 9, 10, 11, 12]	10.37
GO:0018108	peptidyl-tyrosine phosphorylation	9.4E-18	[7, 8, 9]	9.72
GO:0043405	regulation of MAP kinase activity	6.0E-24	[6, 7, 8, 9, 10, 11]	9.62
	transcription initiation from RNA polymerase II			
GO:0006367	promoter	36.0E-15	[7, 8, 9, 10, 11]	9.52
GO:0006352	DNA-templated transcription, initiation	4.3E-15	[6, 7, 8, 9, 10]	9.14
	innate immune response-activating signal			
GO:0002758	transduction	320.0E-15	[4, 6, 7, 8, 9]	9.09
	regulation of sequence-specific DNA binding			
GO:0051090	transcription factor activity	99.0E-18	[3, 6, 7, 8, 9, 10, 11]	9.02
GO:0032147	activation of protein kinase activity	10.0E-21	[7, 8, 9, 10, 11]	8.94
	positive regulation of protein serine/threonine			
GO:0071902	kinase activity	270.0E-15	[7, 8, 9, 10, 11]	8.78
GO:0018105	peptidyl-serine phosphorylation	47.0E-12	[7, 8, 9]	8.71
GO:0070371	ERK1 and ERK2 cascade	60.0E-12	[5, 6, 7, 9, 10]	8.61
	regulation of protein serine/threonine kinase			
GO:0071900	activity	1.0E-24	[7, 8, 9, 10]	8.46
GO:0070372	regulation of ERK1 and ERK2 cascade	620.0E-12	[6, 7, 8, 9, 10, 11]	8.43
GO:0000186	activation of MAPKK activity	2.6E-9	[6, 7, 8, 9, 10, 11, 12]	8.20
GO:0033674	positive regulation of kinase activity	10.0E-24	[5, 6, 7, 8, 9]	8.08
GO:0045860	positive regulation of protein kinase activity	370.0E-24	[6, 7, 8, 9, 10]	7.98
GO:0045859	regulation of protein kinase activity	240.0E-33	[6, 7, 8, 9]	7.95
GO:0043408	regulation of MAPK cascade	120.0E-30	[5, 6, 7, 8, 9, 10]	7.95
GO:0043410	positive regulation of MAPK cascade	9.7E-21	[5, 6, 7, 8, 9, 10, 11]	7.74
GO:0000165	MAPK cascade	14.0E-30	[4, 5, 6, 8, 9]	7.70
GO:0007173	epidermal growth factor receptor signaling pathway	18.0E-12	[8, 9]	7.39
GO:0008286	insulin receptor signaling pathway	390.0E-12	[7, 8, 9]	7.20
GO:0006606	protein import into nucleus	12.0E-9	[4, 5, 6, 7, 8, 9]	7.14
GO:0001525	angiogenesis	15.0E-12	[3, 4, 5, 7, 8, 9, 10]	6.74
GO:0001934	positive regulation of protein phosphorylation	640.0E-27	[6, 7, 8, 9]	6.56
GO:0030098	lymphocyte differentiation	65.0E-9	[4, 6, 7, 8, 9]	6.48
GO:0006816	calcium ion transport	5.6E-9	[8, 9]	6.33
GO:0048514	blood vessel morphogenesis	13.0E-12	[3, 4, 6, 7, 8, 9]	6.22
	positive regulation of transcription from RNA			
GO:0045944	polymerase II promoter	2.5E-21	[6, 7, 8, 9, 10, 11, 12]	6.00
GO:0007411	axon guidance	1.1E-9	[5, 6, 7, 8, 9, 10, 11, 12, 13]	5.37
GO:0006874	cellular calcium ion homeostasis	1.5E-6	[8, 9]	5.34
GO:0045893	positive regulation of transcription, DNA-templated	63.0E-24	[5, 6, 7, 8, 9, 10, 11]	5.30
GO:0048666	neuron development	260.0E-15	[4, 5, 6, 8, 9]	4.44
GO:0031175	neuron projection development	10.0E-12	[4, 5, 6, 7, 9, 10]	4.43
GO:0042110	T cell activation	37.0E-6	[4, 6, 8, 9]	4.34
	cell morphogenesis involved in neuron			
GO:0048667	differentiation	51.0E-9	[5, 6, 7, 8, 9, 10]	4.20

GO Levels = Level of term specificity

% Associated Genes = Percentage of the genes from the SBP predicted targets that were associated with the term, com-pared with all the genes associated with the term.

Table A4.2: The list of 332 GO term that were used in calculating Pathway Distance Similarity (PDS)

GOID	GO Term	Term p-value	Term p-value Corrected with Bonferroni step down	GO Levels	% Associated Genes
GO:0006468	protein phosphorylation	430.0E-264	140.0E-261	[6, 7]	43.38
GO:0071310	cellular response to organic substance	180.0E-261	59.0E-258	(321)	40.12
GO:0044093	positive regulation of molecular function	450.0E-237	140.0E-234	[3]	41.74
GO:1902531	regulation of intracellular signal transduction	51.0E-216	16.0E-213	[4, 5, 6]	41.55
GO:0043085	positive regulation of catalytic activity	620.0E-210	200.0E-207	(321)	42.97
GO:0009967	positive regulation of signal transduction	12.0E-198	4.1E-195	[3, 4, 5, 6]	43.60
GO:0031399	regulation of protein modification process	6.6E-195	2.1E-192	[5, 6, 7]	40.74
GO:0023056	positive regulation of signaling	600.0E-192	190.0E-189	[2, 3, 4]	41.37
GO:0010647	positive regulation of cell communication	700.0E-192	220.0E-189	[3, 4, 5]	41.45
GO:0019220	regulation of phosphate metabolic process	380.0E-186	120.0E-183	[5, 6]	40.31
GO:0042325	regulation of phosphorylation	39.0E-183	12.0E-180	[6, 7]	42.95
GO:0051247	positive regulation of protein metabolic process	15.0E-180	4.8E-177	[4, 5, 6]	41.25
GO:0010941	regulation of cell death	1.1E-177	350.0E-177	[3, 4]	40.84
GO:0001932	regulation of protein phosphorylation	4.9E-174	1.5E-171	[6, 7, 8]	43.45
GO:0043067	regulation of programmed cell death	39.0E-174	12.0E-171	[4, 5]	41.61
GO:0023014	signal transduction by protein phosphorylation	17.0E-171	5.4E-168	[3, 4, 5, 7, 8]	50.05
GO:0032270	positive regulation of cellular protein metabolic process	33.0E-171	10.0E-168	[4, 5, 6, 7]	41.55
GO:0042981	regulation of apoptotic process	430.0E-171	130.0E-168	[5, 6]	41.42
GO:0000165	MAPK cascade	2.4E-165	750.0E-165	[4, 5, 6, 8, 9]	50.27
GO:0031401	positive regulation of protein modification process	480.0E-165	140.0E-162	[5, 6, 7, 8]	44.61
GO:0045937	positive regulation of phosphate metabolic process	41.0E-156	12.0E-153	[5, 6, 7]	45.22
GO:0007167	enzyme linked receptor protein signaling pathway	3.1E-153	960.0E-153	[5, 6]	46.58
GO:0071495	cellular response to endogenous stimulus	390.0E-150	120.0E-147	[3]	42.35
GO:0042327	positive regulation of phosphorylation	5.9E-147	1.8E-144	[6, 7, 8]	46.72
GO:0001934	positive regulation of protein phosphorylation	1.7E-141	530.0E-141	[6, 7, 8, 9]	46.90
GO:0051338	regulation of transferase activity	26.0E-141	7.9E-138	(321)	45.56
GO:0072359	circulatory system development	28.0E-138	8.4E-135	[4, 5]	45.05
GO:0014070	response to organic cyclic compound	60.0E-135	18.0E-132	(321)	45.41
GO:0016477	cell migration	670.0E-135	200.0E-132	[3, 4, 5]	40.09
GO:0022603	regulation of anatomical structure morphogenesis	20.0E-132	6.1E-129	[3, 4]	44.18
GO:0048646	anatomical structure formation involved in morphogenesis	7.7E-129	2.3E-126	[2, 3, 4]	43.47
GO:0098609	cell-cell adhesion	17.0E-129	5.2E-126	[3]	40.66
GO:0051094	positive regulation of developmental process	68.0E-129	20.0E-126	[2, 3, 4]	40.33
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	53.0E-126	15.0E-123	[6, 7]	50.28
GO:1902533	positive regulation of intracellular signal transduction	650.0E-126	190.0E-123	[4, 5, 6, 7]	44.13

GO:0043549	regulation of kinase activity	10.0E-123	3.0E-120	[5, 7, 8]	46.83
GO:1901701	cellular response to oxygen-containing compound	45.0E-123	13.0E-120	(32I)	43.41
GO:0051347	positive regulation of transferase activity	6.8E-120	2.0E-117	[5]	50.52
GO:0033993	response to lipid	22.0E-120	6.6E-117	(32I)	44.10
GO:0060548	negative regulation of cell death	430.0E-117	120.0E-114	[3, 4, 5]	42.84
GO:0051345	positive regulation of hydrolase activity	640.0E-117	180.0E-114	[5]	42.87
GO:0009887	animal organ morphogenesis	66.0E-114	19.0E-111	[3, 4, 5, 6]	41.30
GO:0045859	regulation of protein kinase activity	590.0E-114	170.0E-111	[6, 7, 8, 9]	46.64
GO:0001568	blood vessel development	2.0E-111	570.0E-111	[3, 5, 6, 7, 8]	50.64
GO:0009725	response to hormone	4.2E-111	1.2E-108	[3, 4]	43.50
GO:0048729	tissue morphogenesis	45.0E-111	12.0E-108	[3, 4]	49.17
GO:0072358	cardiovascular system development	940.0E-111	260.0E-108	[4, 5, 6]	49.08
GO:0043069	negative regulation of programmed cell death	12.0E-108	3.5E-105	[4, 5, 6]	43.15
GO:1901698	response to nitrogen compound	90.0E-108	25.0E-105	[3]	42.11
GO:0098602	single organism cell adhesion	160.0E-108	45.0E-105	[2, 3]	44.23
GO:0043066	negative regulation of apoptotic process	1.8E-105	500.0E-105	[5, 6, 7]	43.05
GO:0043068	positive regulation of programmed cell death	12.0E-105	3.3E-102	[4, 5, 6]	49.84
GO:0009611	response to wounding	21.0E-105	5.9E-102	[3]	47.70
GO:0010942	positive regulation of cell death	160.0E-105	45.0E-102	[3, 4, 5]	48.52
GO:0043065	positive regulation of apoptotic process	250.0E-105	69.0E-102	[5, 6, 7]	49.75
GO:0033674	positive regulation of kinase activity	2.5E-102	700.0E-102	[6, 7, 8, 9]	52.15
GO:0043408	regulation of MAPK cascade	74.0E-102	20.0E-99	[5, 6, 7, 8, 9, 10]	45.87
GO:0016337	single organismal cell-cell adhesion	120.0E-102	34.0E-99	[3, 4]	44.58
GO:0048514	blood vessel morphogenesis	260.0E-102	70.0E-99	[3, 4, 6, 7, 8, 9]	51.68
GO:0034097	response to cytokine	820.0E-102	220.0E-99	(32I)	43.05
GO:0001775	cell activation	3.1E-99	830.0E-99	[3]	40.50
GO:0051270	regulation of cellular component movement	4.7E-99	1.2E-96	[3, 4]	42.75
GO:0045597	positive regulation of cell differentiation	530.0E-99	140.0E-96	[3, 4, 5, 6]	41.50
GO:0008284	positive regulation of cell proliferation	790.0E-99	210.0E-96	[3, 4, 5]	41.38
GO:0010243	response to organonitrogen compound	2.2E-96	580.0E-96	[3, 4]	42.58
GO:0018108	peptidyl-tyrosine phosphorylation	3.3E-96	880.0E-96	[7, 8, 9]	58.35
GO:0071407	cellular response to organic cyclic compound	7.8E-96	2.0E-93	[5]	50.27
GO:0045860	positive regulation of protein kinase activity	800.0E-96	210.0E-93	[7, 8, 9, 10]	51.91
GO:2000145	regulation of cell motility	2.4E-93	630.0E-93	[3, 4, 5]	43.13
GO:0030334	regulation of cell migration	8.6E-93	2.2E-90	[4, 5, 6]	44.24
GO:0002009	morphogenesis of an epithelium	90.0E-93	23.0E-90	[4, 5]	49.01
GO:0040012	regulation of locomotion	210.0E-93	56.0E-90	[2, 3]	41.41
GO:0001525	angiogenesis	220.0E-93	58.0E-90	[3, 4, 5, 7, 8, 9, 10]	53.42
GO:0071396	cellular response to lipid	4.1E-90	1.0E-87	[5]	49.17
GO:0042060	wound healing	73.0E-90	18.0E-87	(32I)	48.30
GO:0032870	cellular response to hormone stimulus	240.0E-90	62.0E-87	[4, 5]	45.30
GO:0030155	regulation of cell adhesion	280.0E-90	71.0E-87	[2, 3]	44.25
GO:0070848	response to growth factor	300.0E-90	76.0E-87	(32I)	44.64
GO:0071363	cellular response to growth factor stimulus	4.5E-87	1.1E-84	[5]	45.19
GO:1905114	cell surface receptor signaling pathway involved in cell-cell signaling	130.0E-87	32.0E-84	[4, 5, 6]	47.19

GO:0071900	regulation of protein serine/threonine kinase activity	200.0E-87	50.0E-84	[7, 8, 9, 10]	50.10
GO:0031347	regulation of defense response	310.0E-87	76.0E-84	[4, 5]	43.35
GO:0043410	positive regulation of MAPK cascade	1.8E-84	460.0E-84	[5, 6, 7, 8, 9, 10, 11]	49.12
GO:0071345	cellular response to cytokine stimulus	3.5E-84	850.0E-84	[5]	42.45
GO:0016055	Wnt signaling pathway	700.0E-84	170.0E-81	[5, 6, 7]	48.62
GO:0019221	cytokine-mediated signaling pathway	22.0E-81	5.5E-78	[5, 6]	46.05
GO:1901699	cellular response to nitrogen compound	510.0E-78	120.0E-75	(321)	44.60
GO:2000027	regulation of organ morphogenesis	280.0E-75	68.0E-72	[4, 5, 6, 7]	60.36
GO:0043087	regulation of GTPase activity	28.0E-72	6.7E-69	[5]	40.14
GO:0043405	regulation of MAP kinase activity	140.0E-72	33.0E-69	[6, 7, 8, 9, 10, 11]	53.76
GO:0032147	activation of protein kinase activity	130.0E-72	33.0E-69	[8, 9, 10, 11]	56.05
GO:0071417	cellular response to organonitrogen compound	660.0E-72	150.0E-69	[4, 5]	45.98
GO:0043547	positive regulation of GTPase activity	23.0E-69	5.4E-66	(322)	40.76
GO:0070482	response to oxygen levels	25.0E-69	5.8E-66	[3]	54.46
GO:0046649	lymphocyte activation	100.0E-69	24.0E-66	[3, 5]	40.24
GO:0060070	canonical Wnt signaling pathway	200.0E-69	46.0E-66	[6, 7, 8]	55.34
GO:0007159	leukocyte cell-cell adhesion	320.0E-69	75.0E-66	[4, 5]	44.66
GO:0002764	immune response-regulating signaling pathway	530.0E-69	120.0E-66	[4, 5]	42.51
GO:0051272	positive regulation of cellular component movement	1.0E-66	230.0E-66	[3, 4, 5]	47.30
GO:0045088	regulation of innate immune response	7.2E-66	1.6E-63	[4, 5, 6]	51.10
GO:0051056	regulation of small GTPase mediated signal transduction	14.0E-66	3.2E-63	[5, 6, 7]	53.73
GO:0050878	regulation of body fluid levels	14.0E-66	3.2E-63	[3]	43.69
GO:0031400	negative regulation of protein modification process	20.0E-66	4.7E-63	[5, 6, 7, 8]	41.98
GO:0040017	positive regulation of locomotion	130.0E-66	31.0E-63	[2, 3, 4]	45.81
GO:0097190	apoptotic signaling pathway	170.0E-66	38.0E-63	[4, 5, 6]	40.57
GO:0006935	chemotaxis	340.0E-66	76.0E-63	[3, 4]	41.60
GO:0061061	muscle structure development	370.0E-66	82.0E-63	[3]	41.05
GO:0036293	response to decreased oxygen levels	500.0E-66	110.0E-63	(321)	54.43
GO:0007507	heart development	1.0E-63	240.0E-63	[3, 4, 5, 6]	42.75
GO:0050852	T cell receptor signaling pathway	8.5E-63	1.8E-60	[6, 8, 9, 10]	66.84
GO:0001666	response to hypoxia	11.0E-63	2.5E-60	[3, 5]	54.55
GO:0048545	response to steroid hormone	45.0E-63	9.7E-60	[4, 5]	47.55
GO:0070486	leukocyte aggregation	750.0E-63	160.0E-60	[5, 6]	44.17
GO:0018105	peptidyl-serine phosphorylation	1.2E-60	250.0E-60	[7, 8, 9]	56.39
GO:0048732	gland development	1.8E-60	390.0E-60	[3, 4, 5, 6]	44.54
GO:0002521	leukocyte differentiation	2.0E-60	420.0E-60	[5, 6, 7, 8]	43.83
GO:0042110	T cell activation	3.2E-60	690.0E-60	[4, 6, 8, 9]	44.28
GO:0002768	immune response-regulating cell surface receptor signaling pathway	11.0E-60	2.5E-57	[5, 6]	45.17
GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	45.0E-60	9.4E-57	[3, 6, 7, 8, 9, 10, 11]	47.67
GO:0046777	protein autophosphorylation	68.0E-60	14.0E-57	[7, 8]	57.49
GO:0002757	immune response-activating signal transduction	78.0E-60	16.0E-57	[3, 5, 6, 7]	41.35
GO:0018209	peptidyl-serine modification	100.0E-60	21.0E-57	[7, 8]	54.48
GO:0070997	neuron death	150.0E-60	30.0E-57	(321)	52.10
GO:1901652	response to peptide	390.0E-60	81.0E-57	[4, 5]	43.97
GO:0045785	positive regulation of cell adhesion	1.0E-57	200.0E-57	[2, 3, 4]	45.89

GO:0050851	antigen receptor-mediated signaling pathway	1.1E-57	220.0E-57	[5, 7, 8, 9]	56.57
GO:0031349	positive regulation of defense response	1.3E-57	280.0E-57	[3, 4, 5, 6]	45.99
GO:1901342	regulation of vasculature development	7.9E-57	1.6E-54	[4, 5, 6, 7, 8]	55.64
GO:0030111	regulation of Wnt signaling pathway	8.6E-57	1.7E-54	[4, 5, 6, 7, 8]	49.85
GO:0060537	muscle tissue development	12.0E-57	2.5E-54	(321)	47.00
GO:0030198	extracellular matrix organization	20.0E-57	4.1E-54	(321)	48.99
GO:0007599	hemostasis	33.0E-57	6.5E-54	(321)	47.68
GO:0035239	tube morphogenesis	41.0E-57	8.1E-54	[3, 4, 5]	47.44
GO:0043434	response to peptide hormone	100.0E-57	20.0E-54	[4, 5, 6]	44.81
GO:0050817	coagulation	190.0E-57	38.0E-54	[3]	47.41
GO:1901214	regulation of neuron death	260.0E-57	51.0E-54	[4, 5]	53.68
GO:0050900	leukocyte migration	740.0E-57	140.0E-54	[2, 4, 5, 6]	45.82
GO:0022407	regulation of cell-cell adhesion	840.0E-57	160.0E-54	[3, 4, 5]	45.12
GO:0050673	epithelial cell proliferation	2.5E-54	480.0E-54	[3]	46.21
GO:0071902	positive regulation of protein serine/threonine kinase activity	2.7E-54	510.0E-54	[8, 9, 10, 11]	50.00
GO:1905330	regulation of morphogenesis of an epithelium	42.0E-54	8.0E-51	[4, 5, 6]	61.58
GO:0014706	striated muscle tissue development	66.0E-54	12.0E-51	[5]	46.47
GO:2001233	regulation of apoptotic signaling pathway	68.0E-54	12.0E-51	[4, 5, 6, 7]	44.23
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	89.0E-54	16.0E-51	[5]	66.25
GO:0060562	epithelial tube morphogenesis	150.0E-54	28.0E-51	[4, 5, 6]	48.20
GO:0001667	ameboid-type cell migration	170.0E-54	32.0E-51	[4, 5, 6]	47.16
GO:0045862	positive regulation of proteolysis	380.0E-54	69.0E-51	[5, 6, 7, 8]	46.20
GO:0007265	Ras protein signal transduction	630.0E-54	110.0E-51	[6, 7]	47.01
GO:0045765	regulation of angiogenesis	5.5E-51	1.0E-48	[4, 5, 6, 7, 8, 9, 10, 11]	55.36
GO:0034330	cell junction organization	6.9E-51	1.2E-48	[3]	52.04
GO:0048017	inositol lipid-mediated signaling	17.0E-51	3.0E-48	[5, 6]	56.62
GO:0002429	immune response-activating cell surface receptor signaling pathway	21.0E-51	3.7E-48	[4, 6, 7, 8]	43.48
GO:0031589	cell-substrate adhesion	45.0E-51	8.0E-48	[3]	47.43
GO:0048015	phosphatidylinositol-mediated signaling	78.0E-51	13.0E-48	[6, 7]	56.74
GO:0001763	morphogenesis of a branching structure	87.0E-51	15.0E-48	[3, 4]	57.42
GO:0043406	positive regulation of MAP kinase activity	350.0E-51	61.0E-48	[6, 7, 8, 9, 10, 11, 12]	54.51
GO:0050730	regulation of peptidyl-tyrosine phosphorylation	350.0E-51	61.0E-48	[7, 8, 9, 10]	54.51
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	350.0E-51	61.0E-48	(321)	54.51
GO:0061138	morphogenesis of a branching epithelium	430.0E-51	74.0E-48	[4, 5, 6]	58.67
GO:0014065	phosphatidylinositol 3-kinase signaling	1.3E-48	230.0E-48	[7, 8]	61.99
GO:0002223	stimulatory C-type lectin receptor signaling pathway	1.7E-48	290.0E-48	[6, 8, 9, 10, 11]	75.93
GO:0051348	negative regulation of transferase activity	1.7E-48	290.0E-48	[5]	45.05
GO:0014066	regulation of phosphatidylinositol 3-kinase signaling	4.9E-48	830.0E-48	[5, 6, 7, 8, 9]	65.33
GO:0003012	muscle system process	5.0E-48	840.0E-48	[3]	43.07
GO:0002218	activation of innate immune response	5.3E-48	880.0E-48	[3, 5, 6, 7, 8]	51.75
GO:0060828	regulation of canonical Wnt signaling pathway	14.0E-48	2.3E-45	[5, 6, 7, 8, 9]	51.98
GO:0051402	neuron apoptotic process	14.0E-48	2.4E-45	[5, 6]	54.42

GO:0097191	extrinsic apoptotic signaling pathway	66.0E-48	10.0E-45	[5, 6, 7]	52.02
GO:0071383	cellular response to steroid hormone stimulus	80.0E-48	13.0E-45	[5, 6]	51.37
GO:1903037	regulation of leukocyte cell-cell adhesion	100.0E-48	16.0E-45	[4, 5, 6]	45.83
GO:0001933	negative regulation of protein phosphorylation	160.0E-48	26.0E-45	[6, 7, 8, 9]	42.82
GO:0038093	Fc receptor signaling pathway	250.0E-48	40.0E-45	[6, 7]	49.64
GO:0050678	regulation of epithelial cell proliferation	280.0E-48	44.0E-45	[4, 5]	46.56
GO:0002237	response to molecule of bacterial origin	300.0E-48	47.0E-45	[4, 6]	45.32
GO:0045089	positive regulation of innate immune response	340.0E-48	53.0E-45	[4, 5, 6, 7]	47.83
GO:0042326	negative regulation of phosphorylation	560.0E-48	87.0E-45	[6, 7, 8]	41.15
GO:1901653	cellular response to peptide	1.6E-45	250.0E-45	[5, 6]	45.62
GO:0032496	response to lipopolysaccharide	2.2E-45	340.0E-45	[4, 5, 7]	45.73
GO:0035567	non-canonical Wnt signaling pathway	2.4E-45	370.0E-45	[6, 7, 8]	61.88
GO:0048738	cardiac muscle tissue development	2.6E-45	400.0E-45	[4, 5, 6, 7]	56.63
GO:0043401	steroid hormone mediated signaling pathway	2.8E-45	420.0E-45	[5, 6, 7]	56.99
GO:0030098	lymphocyte differentiation	3.4E-45	510.0E-45	[4, 6, 7, 8, 9]	46.23
GO:0090130	tissue migration	14.0E-45	2.1E-42	[3]	50.39
GO:0070371	ERK1 and ERK2 cascade	22.0E-45	3.3E-42	[5, 6, 7, 9, 10]	48.39
GO:0009755	hormone-mediated signaling pathway	22.0E-45	3.3E-42	[4, 5, 6]	51.68
GO:0007517	muscle organ development	75.0E-45	10.0E-42	[4, 5, 6]	42.52
GO:0031098	stress-activated protein kinase signaling cascade	80.0E-45	11.0E-42	[4, 5, 6]	49.24
GO:0050863	regulation of T cell activation	79.0E-45	11.0E-42	[5, 6, 7, 9, 10]	45.48
GO:0030178	negative regulation of Wnt signaling pathway	100.0E-45	14.0E-42	[4, 5, 6, 7, 8, 9]	54.03
GO:0071375	cellular response to peptide hormone stimulus	120.0E-45	17.0E-42	[5, 6, 7]	45.98
GO:0001503	ossification	170.0E-45	23.0E-42	[3]	42.30
GO:0051403	stress-activated MAPK cascade	220.0E-45	30.0E-42	[5, 6, 7, 9, 10]	49.42
GO:0060326	cell chemotaxis	390.0E-45	54.0E-42	[4, 5, 6]	48.18
GO:0048754	branching morphogenesis of an epithelial tube	560.0E-45	77.0E-42	[4, 5, 6, 7]	60.25
GO:0030177	positive regulation of Wnt signaling pathway	560.0E-45	77.0E-42	[4, 5, 6, 7, 8, 9]	60.25
GO:0030168	platelet activation	590.0E-45	80.0E-42	[4, 5, 6]	58.38
GO:0043523	regulation of neuron apoptotic process	610.0E-45	83.0E-42	[5, 6, 7]	54.73
GO:0090132	epithelium migration	970.0E-45	130.0E-42	(32I)	50.00
GO:0050867	positive regulation of cell activation	3.0E-42	400.0E-42	[3, 4, 5]	43.60
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	4.8E-42	640.0E-42	[7, 8, 9, 10, 11]	56.98
GO:0090090	negative regulation of canonical Wnt signaling pathway	10.0E-42	1.4E-39	[5, 6, 7, 8, 9, 10]	57.47
GO:0042176	regulation of protein catabolic process	12.0E-42	1.6E-39	[4, 5, 6]	41.62
GO:0022409	positive regulation of cell-cell adhesion	19.0E-42	2.4E-39	[3, 4, 5, 6]	48.11
GO:0038095	Fc-epsilon receptor signaling pathway	27.0E-42	3.4E-39	[7, 8]	53.14
GO:0002696	positive regulation of leukocyte activation	27.0E-42	3.4E-39	[3, 4, 5, 6]	43.71
GO:0045216	cell-cell junction organization	47.0E-42	6.0E-39	(32I)	50.00
GO:0042692	muscle cell differentiation	81.0E-42	10.0E-39	[4, 5]	41.60
GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity	88.0E-42	10.0E-39	[2, 3, 4, 7, 8, 9, 10, 11, 12]	49.19

GO:0070372	regulation of ERK1 and ERK2 cascade	90.0E-42	11.0E-39	[6, 7, 8, 9, 10, 11]	48.08
GO:0060759	regulation of response to cytokine stimulus	100.0E-42	12.0E-39	[3, 4, 5]	60.13
GO:0031329	regulation of cellular catabolic process	280.0E-42	34.0E-39	[4, 5]	40.77
GO:0001655	urogenital system development	370.0E-42	45.0E-39	[4, 5]	42.73
GO:0071453	cellular response to oxygen levels	470.0E-42	56.0E-39	(321)	58.86
GO:0007160	cell-matrix adhesion	1.1E-39	130.0E-39	(321)	51.90
GO:0034612	response to tumor necrosis factor	1.1E-39	130.0E-39	[5]	45.94
GO:1903039	positive regulation of leukocyte cell-cell adhesion	1.4E-39	160.0E-39	[4, 5, 6, 7]	49.57
GO:1904018	positive regulation of vasculature development	2.0E-39	230.0E-39	[3, 4, 5, 6, 7, 8, 9]	59.60
GO:0051251	positive regulation of lymphocyte activation	4.1E-39	470.0E-39	[4, 5, 6, 7]	44.16
GO:0003007	heart morphogenesis	5.6E-39	640.0E-39	[4, 5, 6, 7]	48.75
GO:0034329	cell junction assembly	6.7E-39	750.0E-39	(321)	51.17
GO:1901215	negative regulation of neuron death	7.9E-39	880.0E-39	[4, 5, 6]	54.01
GO:0001959	regulation of cytokine-mediated signaling pathway	8.9E-39	980.0E-39	[4, 5, 6, 7]	59.86
GO:0060071	Wnt signaling pathway, planar cell polarity pathway	10.0E-39	1.1E-36	[6, 7, 8, 9]	66.38
GO:0071356	cellular response to tumor necrosis factor	14.0E-39	1.5E-36	(322)	46.77
GO:0043542	endothelial cell migration	58.0E-39	6.2E-36	[6, 7, 8]	55.56
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	120.0E-39	12.0E-36	[6, 7]	42.09
GO:0043281	regulation of cysteine-type endopeptidase activity involved in apoptotic process	130.0E-39	14.0E-36	[6, 7, 8, 9, 10]	50.23
GO:0050870	positive regulation of T cell activation	200.0E-39	21.0E-36	[5, 6, 7, 8, 9, 10, 11]	49.12
GO:0045766	positive regulation of angiogenesis	320.0E-39	32.0E-36	[4, 5, 6, 7, 8, 9, 10, 11, 12]	61.19
GO:2001236	regulation of extrinsic apoptotic signaling pathway	320.0E-39	32.0E-36	[5, 6, 7, 8]	53.89
GO:0001738	morphogenesis of a polarized epithelium	360.0E-39	37.0E-36	[5, 6]	59.44
GO:0009896	positive regulation of catabolic process	420.0E-39	42.0E-36	[3, 4, 5]	41.52
GO:0036294	cellular response to decreased oxygen levels	780.0E-39	77.0E-36	[5]	58.50
GO:0051098	regulation of binding	840.0E-39	82.0E-36	[3]	43.04
GO:0001736	establishment of planar polarity	1.3E-36	130.0E-36	[4, 5, 6, 7]	61.54
GO:0032943	mononuclear cell proliferation	1.8E-36	170.0E-36	(321)	44.77
GO:0070661	leukocyte proliferation	2.9E-36	270.0E-36	[3]	43.69
GO:0034332	adherens junction organization	5.6E-36	520.0E-36	[5]	61.90
GO:0071456	cellular response to hypoxia	7.2E-36	670.0E-36	[4, 6]	58.87
GO:0030099	myeloid cell differentiation	9.1E-36	840.0E-36	[5, 6, 7, 8]	40.75
GO:0071559	response to transforming growth factor beta	13.0E-36	1.1E-33	[3, 5]	48.02
GO:2001234	negative regulation of apoptotic signaling pathway	15.0E-36	1.4E-33	[4, 5, 6, 7, 8]	47.06
GO:0090263	positive regulation of canonical Wnt signaling pathway	25.0E-36	2.3E-33	[5, 6, 7, 8, 9, 10]	61.60
GO:0031331	positive regulation of cellular catabolic process	31.0E-36	2.8E-33	[4, 5, 6]	44.20
GO:2000116	regulation of cysteine-type endopeptidase activity	42.0E-36	3.6E-33	[7, 8, 9]	46.67
GO:0070302	regulation of stress-activated protein kinase signaling cascade	59.0E-36	5.0E-33	[4, 5, 6, 7]	49.28
GO:0007259	JAK-STAT cascade	87.0E-36	7.4E-33	[6, 7]	51.61
GO:0010565	regulation of cellular ketone metabolic process	110.0E-36	9.4E-33	[4, 5]	53.53

GO:0006936	muscle contraction	250.0E-36	21.0E-33	(321)	40.79
GO:0000187	activation of MAPK activity intrinsic apoptotic signaling pathway	390.0E-36	32.0E-33	[7, 8, 9, 10, 11, 12, 13]	55.13
GO:0097193	regulation of protein localization to nucleus	570.0E-36	46.0E-33	[5, 6, 7]	42.28
GO:1900180	regulation of muscle system process	910.0E-36	73.0E-33	[5, 6, 7]	46.55
GO:0090257	regulation of proteolysis involved in cellular protein catabolic process	950.0E-36	75.0E-33	[4, 5]	49.75
GO:1903050	regulation of cellular response to growth factor stimulus	1.4E-33	110.0E-33	[6, 7, 8]	46.35
GO:0090287	regulation of hemopoiesis	2.8E-33	220.0E-33	[3, 4, 6]	44.98
GO:1903706	positive regulation of epithelial cell proliferation	7.1E-33	540.0E-33	[3, 4, 5, 6, 7, 8]	40.43
GO:0050679	regulation of ossification	8.9E-33	660.0E-33	[4, 5, 6]	51.72
GO:0030278	Rho protein signal transduction	9.6E-33	710.0E-33	[3, 4]	49.74
GO:0007266	regulation of nucleocytoplasmic transport	13.0E-33	950.0E-33	[7, 8]	51.12
GO:0046822	regulation of cellular protein catabolic process	38.0E-33	2.7E-30	[5, 6, 7]	45.85
GO:1903362	positive regulation of protein catabolic process	64.0E-33	4.5E-30	[5, 6, 7]	43.75
GO:0045732	regulation of protein modification by small protein conjugation or removal	77.0E-33	5.4E-30	[4, 5, 6, 7]	43.18
GO:1903320	transcription initiation from RNA polymerase II promoter	160.0E-33	11.0E-30	[6, 7, 8]	41.75
GO:0006367	osteoblast differentiation	210.0E-33	14.0E-30	[7, 8, 9, 10, 11]	49.73
GO:0001649	regulation of protein import	260.0E-33	17.0E-30	[4, 5]	46.51
GO:1904589	response to acid chemical	280.0E-33	18.0E-30	[5, 6, 7]	49.20
GO:0001101	connective tissue development	1.1E-30	75.0E-30	[3]	41.03
GO:0061448	regulation of epithelial cell migration	1.3E-30	84.0E-30	(321)	43.55
GO:0010632	regulation of protein import into nucleus	1.6E-30	100.0E-30	[3, 4, 5, 6, 7, 8]	49.44
GO:0042306	T cell differentiation	2.2E-30	140.0E-30	[6, 7, 8, 9, 10]	48.91
GO:0030217	tumor necrosis factor-mediated signaling pathway	2.3E-30	140.0E-30	[5, 7, 8, 9, 10]	45.83
GO:0033209	positive regulation of proteolysis involved in cellular protein catabolic process	3.4E-30	200.0E-30	[6, 7]	51.55
GO:1903052	positive regulation of cellular protein catabolic process	4.5E-30	270.0E-30	[6, 7, 8, 9]	49.71
GO:1903364	striated muscle cell differentiation	4.9E-30	280.0E-30	[5, 6, 7, 8]	48.15
GO:0051146	negative regulation of kinase activity	5.4E-30	300.0E-30	[5, 6]	42.47
GO:0033673	renal system development	11.0E-30	620.0E-30	[6, 7, 8, 9]	42.41
GO:0072001	DNA-templated transcription, initiation	11.0E-30	620.0E-30	[4, 5, 6]	40.00
GO:0006352	gliogenesis	14.0E-30	750.0E-30	[6, 7, 8, 9, 10]	44.16
GO:0042063	transforming growth factor beta receptor signaling pathway	15.0E-30	780.0E-30	[6, 7]	42.80
GO:0007179	protein import into nucleus	17.0E-30	910.0E-30	[5, 7, 8]	48.62
GO:0006606	kidney development	23.0E-30	1.1E-27	[5, 6, 7, 8, 9]	40.14
GO:0001822	regulation of Ras protein signal transduction	32.0E-30	1.6E-27	[3, 4, 5, 6, 7]	40.64
GO:0046578	regulation of protein ubiquitination	36.0E-30	1.7E-27	[6, 7, 8]	44.59
GO:0031396	negative regulation of protein kinase activity	39.0E-30	1.8E-27	[7, 8, 9, 10]	41.92
GO:0006469	mesenchyme development	57.0E-30	2.6E-27	[7, 8, 9, 10]	43.28
GO:0060485	positive regulation of protein modification by small protein conjugation or removal	78.0E-30	3.5E-27	[4, 5, 6]	42.13
GO:1903322		140.0E-30	6.3E-27	[6, 7, 8, 9]	45.81

GO:0050920	regulation of chemotaxis	190.0E-30	8.3E-27	[3, 4, 5]	46.00
GO:0032944	regulation of mononuclear cell proliferation	190.0E-30	8.3E-27	[5, 6]	45.02
GO:2000377	regulation of reactive oxygen species metabolic process	360.0E-30	15.0E-27	[4, 5]	49.40
GO:0070663	regulation of leukocyte proliferation	400.0E-30	16.0E-27	[4, 5]	44.09
GO:0043524	negative regulation of neuron apoptotic process	410.0E-30	16.0E-27	[5, 6, 7, 8]	52.05
GO:0072073	kidney epithelium development	1.5E-27	59.0E-27	[4, 5, 6, 7, 8]	50.32
GO:0010810	regulation of cell-substrate adhesion	2.2E-27	83.0E-27	[3, 4]	45.45
GO:0042113	B cell activation	3.5E-27	120.0E-27	[4, 6]	41.27
GO:0071216	cellular response to biotic stimulus	4.0E-27	140.0E-27	[3]	45.83
GO:0002263	cell activation involved in immune response	5.6E-27	190.0E-27	[3, 4]	42.61
GO:0002573	myeloid leukocyte differentiation	8.8E-27	290.0E-27	[6, 7, 8, 9]	44.78
GO:0007254	JNK cascade	8.8E-27	290.0E-27	[6, 7, 8, 10, 11]	44.78
GO:0070374	positive regulation of ERK1 and ERK2 cascade	9.7E-27	310.0E-27	[6, 7, 8, 9, 10, 11, 12]	46.45
GO:0010952	positive regulation of peptidase activity	10.0E-27	330.0E-27	[6, 7, 8, 9]	48.77
GO:2001235	positive regulation of apoptotic signaling pathway	11.0E-27	350.0E-27	[4, 5, 6, 7, 8]	45.99
GO:0072593	reactive oxygen species metabolic process	23.0E-27	690.0E-27	[3]	40.47
GO:0046425	regulation of JAK-STAT cascade	29.0E-27	810.0E-27	[6, 7, 8]	49.04
GO:1903034	regulation of response to wounding	38.0E-27	1.0E-24	[4, 5]	48.45
GO:0031398	positive regulation of protein ubiquitination	40.0E-27	1.0E-24	[7, 8, 9, 10, 11]	45.70
GO:1902105	regulation of leukocyte differentiation	47.0E-27	1.1E-24	[4, 5, 6, 7, 8, 9]	40.89
GO:0043393	regulation of protein binding	53.0E-27	1.2E-24	(32I)	45.90
GO:0006140	regulation of nucleotide metabolic process	56.0E-27	1.3E-24	[5, 6, 7, 8]	42.11
GO:0030595	leukocyte chemotaxis	180.0E-27	4.1E-24	[3, 5, 6, 7]	43.27
GO:0051216	cartilage development	240.0E-27	5.2E-24	[4, 5, 6]	44.74
GO:0042180	cellular ketone metabolic process	430.0E-27	8.2E-24	[3, 4]	40.66
GO:0043122	regulation of I-kappaB kinase/NF-kappaB signaling	430.0E-27	8.2E-24	[5, 6, 7]	40.66
GO:2000146	negative regulation of cell motility	420.0E-27	8.4E-24	[3, 4, 5, 6]	40.93
GO:0048762	mesenchymal cell differentiation	1.0E-24	16.0E-24	[5, 6, 7]	43.65
GO:0071219	cellular response to molecule of bacterial origin	940.0E-27	16.0E-24	[4, 5, 7]	46.20
GO:0071229	cellular response to acid chemical	1.2E-24	16.0E-24	(32I)	46.43
GO:0030336	negative regulation of cell migration	1.1E-24	16.0E-24	[4, 5, 6, 7]	41.52
GO:0003205	cardiac chamber development	1.3E-24	16.0E-24	[3, 4, 5, 6, 7]	47.20
GO:0042098	T cell proliferation	1.1E-24	17.0E-24	[5, 6, 7, 9, 10]	44.62
GO:0071222	cellular response to lipopolysaccharide	1.0E-24	17.0E-24	[5, 6, 8]	46.95
GO:0046328	regulation of JNK cascade	2.0E-24	22.0E-24	[6, 7, 8, 9, 10, 11, 12]	46.15
GO:0097529	myeloid leukocyte migration	2.8E-24	28.0E-24	[3, 5, 6, 7]	45.20
GO:0002685	regulation of leukocyte migration	5.7E-24	51.0E-24	[3, 4, 5, 6, 7]	46.34
GO:0060541	respiratory system development	6.9E-24	55.0E-24	[4, 5]	41.78
GO:0051147	regulation of muscle cell differentiation	8.1E-24	57.0E-24	[4, 5, 6]	45.35
GO:1900542	regulation of purine nucleotide metabolic process	10.0E-24	64.0E-24	[6, 7, 8, 9]	41.28
GO:0045637	regulation of myeloid cell differentiation	62.0E-24	310.0E-24	[4, 5, 6, 7, 8, 9]	43.48
GO:0031960	response to corticosteroid	340.0E-24	1.0E-21	[5, 6]	43.93
GO:0030323	respiratory tube development	330.0E-24	1.3E-21	[3, 4, 5, 6]	42.55

GO:0010001	glial cell differentiation	1.5E-21	3.1E-21	[5, 7, 8]	42.02
GO:0032869	cellular response to insulin stimulus	5.4E-21	5.4E-21	[6, 7, 8]	40.70

Table A4.3: The list of additional thirteen targets from literature, which ginsenoside compounds are found to be implicated in the modulation of the targets.

Gene ID	Protein name	SBP compound	Ref.
BAX	Apoptosis regulator BAX	ginsenoside Rd	(323)
BCL2	Apoptosis regulator Bcl-2	ginsenoside Rd	(323)
CASP3	Caspase-3	ginsenoside Rb3	(301)
CASP8	Caspase-8	ginsenoside Rb3	(301)
CASP9	Caspase-9	ginsenoside Rb3	(301)
EGF	Pro-epidermal growth factor	ginsenoside Rb2	(308)
EGFR	Epidermal growth factor receptor	ginsenoside Rb2	(308)
ESR2	Estrogen beta receptor	ginsenoside Rb1	(324)
FN1	Fibronectin	ginsenoside Rb2	(308)
FOS	Proto-oncogene c-Fos	ginsenoside Rc	(325)
FOS	Proto-oncogene c-Fos	ginsenoside Re	(325)
MMP1	Interstitial collagenase	ginsenoside Rb2	(308)
NOS3	Nitric oxide synthase	ginsenoside Re	(326)
NOS3	Nitric oxide synthase	ginsenoside Rg3	(327)
NR3C1	Glucocorticoid receptor	ginsenoside Re	(328)
NR3C1	Glucocorticoid receptor	ginsenoside Rg1	(329)

Table A4.4: The values of SS, TS, PDS and MDS of 231 combinations of SBP.

ID	compound <i>i</i>	compound <i>j</i>	TS	PDS	MDS	SS
108	Ginsenoside Rb3	Cholic acid	0.0061373	0.5104493	2.849	0.00109953
119	Ginsenoside Rc	Cholic acid	0.0055149	0.5713448	2.889	0.00109070
153	Ginsenoside Re	Ginsenoside Rb3	0.0067909	0.4234194	2.646	0.00108676
128	Ginsenoside Rc	Ginsenoside Rb3	0.0095032	0.3102916	2.821	0.00104547
115	Ginsenoside Rb3	Ginsenoside Rb1	0.0097747	0.3023365	2.949	0.00100222
117	Ginsenoside Rb3	17-hydroxyprogesterone	0.0059602	0.4826925	2.876	0.00100025
116	Ginsenoside Rb3	Ginsenoside Rb2	0.0064827	0.4052489	2.635	0.00099714
170	Ginsenoside Rg1	Ginsenoside Rb3	0.0069584	0.4119944	2.944	0.00097363
89	Ginsenoside Rb1	Cholic acid	0.0057864	0.4769261	2.916	0.00094643
112	Ginsenoside Rb3	Deoxycholic acid	0.0060271	0.4404201	2.920	0.00090906
113	Ginsenoside Rb3	11-hydroxyprogesterone	0.0059610	0.5081140	3.333	0.00090866
140	Ginsenoside Rd	Ginsenoside Rb3	0.0093223	0.2742955	2.864	0.00089277
114	Ginsenoside Rb3	Gamabufotalin	0.0058351	0.4490800	2.940	0.00089116
129	Ginsenoside Rc	17-hydroxyprogesterone	0.0053377	0.4900444	2.956	0.00088502
169	Ginsenoside Rg1	Ginsenoside Rb2	0.0033154	0.7187246	2.758	0.00086397
139	Ginsenoside Rd	Ginsenoside Rb2	0.0056794	0.4104020	2.710	0.00086019
107	Ginsenoside Rb3	Bufalin	0.0058343	0.4229923	2.916	0.00084633
171	Ginsenoside Rg1	Ginsenoside Rc	0.0063359	0.3852619	2.911	0.00083851
111	Ginsenoside Rb3	Cinobufagin	0.0059342	0.4222491	2.990	0.00083809
120	Ginsenoside Rc	Chenodeoxycholic acid	0.0054058	0.4518561	2.940	0.00083083
123	Ginsenoside Rc	Deoxycholic acid	0.0054047	0.4511290	2.953	0.00082557
124	Ginsenoside Rc	11-hydroxyprogesterone	0.0053386	0.5110724	3.423	0.00079706
208	Hyodeoxycholic acid	Ginsenoside Rb3	0.0060221	0.3907532	2.972	0.00079171

109	Ginsenoside Rb3	Chenodeoxycholic acid	0.0060282	0.3770864	2.886	0.00078773
209	Hyodeoxycholic acid	Ginsenoside Rc	0.0053997	0.4339484	3.032	0.00077271
127	Ginsenoside Rc	Ginsenoside Rb2	0.0058602	0.3387305	2.578	0.00077006
190	Ginsenoside Rg3	Ginsenoside Rb3	0.0077555	0.2819174	2.851	0.00076696
168	Ginsenoside Rg1	Ginsenoside Rb1	0.0066075	0.3369864	2.956	0.00075337
96	Ginsenoside Rb1	17-hydroxyprogesterone	0.0056093	0.4004635	2.994	0.00075017
131	Ginsenoside Rd	Cholic acid	0.0053340	0.3926011	2.798	0.00074841
29	Borneol	Ginsenoside Rb3	0.0058267	0.3746318	2.940	0.00074237
152	Ginsenoside Re	Ginsenoside Rb2	0.0031479	0.5823789	2.491	0.00073599
155	Ginsenoside Re	Ginsenoside Rd	0.0059876	0.3211465	2.675	0.00071884
90	Ginsenoside Rb1	Chenodeoxycholic acid	0.0056773	0.3744867	2.973	0.00071505
225	Muscone	Ginsenoside Rb3	0.0058999	0.3404798	2.836	0.00070839
93	Ginsenoside Rb1	Deoxycholic acid	0.0056762	0.3695216	2.967	0.00070701
94	Ginsenoside Rb1	11-hydroxyprogesterone	0.0056101	0.4240120	3.433	0.00069281
105	Ginsenoside Rb2	Ginsenoside Rb1	0.0061318	0.2894868	2.667	0.00066565
206	Hyodeoxycholic acid	Ginsenoside Rb1	0.0056712	0.3600544	3.069	0.00066525
122	Ginsenoside Rc	Cinobufagin	0.0053118	0.3862229	3.089	0.00066408
142	Ginsenoside Rd	17-hydroxyprogesterone	0.0051569	0.3698512	2.877	0.00066284
61	Cinnamaldehyde	Ginsenoside Rb3	0.0056234	0.3413805	2.959	0.00064868
118	Ginsenoside Rc	Bufalin	0.0052119	0.3730060	3.056	0.00063624
125	Ginsenoside Rc	Gamabufotalin	0.0052127	0.3735706	3.065	0.00063537
11	Benzyl benzoate	Ginsenoside Rb3	0.0056860	0.3370369	3.026	0.00063323
136	Ginsenoside Rd	11-hydroxyprogesterone	0.0051577	0.4028853	3.310	0.00062769
98	Ginsenoside Rb2	Cholic acid	0.0024944	0.6611088	2.639	0.00062490
189	Ginsenoside Rg3	Ginsenoside Rb2	0.0041126	0.4024809	2.649	0.00062477
135	Ginsenoside Rd	Deoxycholic acid	0.0052238	0.3369640	2.855	0.00061650
110	Ginsenoside Rb3	Cinnamic acid	0.0055391	0.3304779	2.996	0.00061099
172	Ginsenoside Rg1	Ginsenoside Rd	0.0061551	0.2827583	2.879	0.00060449
141	Ginsenoside Rd	Ginsenoside Rc	0.0086999	0.1922611	2.896	0.00057767
30	Borneol	Ginsenoside Rc	0.0052042	0.3397520	3.086	0.00057302
95	Ginsenoside Rb1	Gamabufotalin	0.0054842	0.3215150	3.116	0.00056592
210	Hyodeoxycholic acid	Ginsenoside Rd	0.0052188	0.3172761	2.932	0.00056466
223	Muscone	Ginsenoside Rb1	0.0055490	0.3035850	3.013	0.00055920
88	Ginsenoside Rb1	Bufalin	0.0054834	0.3104770	3.102	0.00054890
226	Muscone	Ginsenoside Rc	0.0052775	0.3032066	2.942	0.00054397
27	Borneol	Ginsenoside Rb1	0.0054758	0.3055238	3.120	0.00053629
132	Ginsenoside Rd	Chenodeoxycholic acid	0.0052249	0.2926457	2.852	0.00053616
92	Ginsenoside Rb1	Cinobufagin	0.0055833	0.2983934	3.131	0.00053211
62	Cinnamaldehyde	Ginsenoside Rc	0.0050010	0.3224239	3.050	0.00052873
59	Cinnamaldehyde	Ginsenoside Rb1	0.0052725	0.3083327	3.115	0.00052195
151	Ginsenoside Re	Ginsenoside Rb1	0.0064400	0.2047906	2.667	0.00049457
138	Ginsenoside Rd	Ginsenoside Rb1	0.0089714	0.1550928	2.910	0.00047807
9	Benzyl benzoate	Ginsenoside Rb1	0.0053351	0.2793350	3.131	0.00047601
12	Benzyl benzoate	Ginsenoside Rc	0.0050636	0.2850652	3.077	0.00046912
102	Ginsenoside Rb2	Deoxycholic acid	0.0023841	0.5240455	2.710	0.00046103
121	Ginsenoside Rc	Cinnamic acid	0.0049167	0.2897426	3.094	0.00046049
106	Ginsenoside Rb2	17-hydroxyprogesterone	0.0023172	0.5427370	2.738	0.00045941
180	Ginsenoside Rg3	Cholic acid	0.0037672	0.3401296	2.791	0.00045910

99	Ginsenoside Rb2	Chenodeoxycholic acid	0.0023853	0.5176688	2.700	0.00045732
91	Ginsenoside Rb1	Cinnamic acid	0.0051882	0.2761886	3.158	0.00045381
137	Ginsenosde Rd	Gamabufotalin	0.0050318	0.2695963	3.000	0.00045219
103	Ginsenoside Rb2	11-hydroxyprogesterone	0.0023181	0.6164160	3.160	0.00045214
134	Ginsenosde Rd	Cinobufagin	0.0051310	0.2669244	3.039	0.00045068
173	Ginsenoside Rg1	Ginsenoside Re	0.0036236	0.3389415	2.731	0.00044976
167	Ginsenoside Rg1	Gamabufotalin	0.0026679	0.4949464	2.955	0.00044692
154	Ginsenoside Re	Ginsenoside Rc	0.0061685	0.1912693	2.658	0.00044390
197	Ginsenoside Rg3	17-hydroxyprogesterone	0.0035900	0.3469515	2.839	0.00043866
130	Ginsenosde Rd	Bufalin	0.0050311	0.2596079	2.982	0.00043800
207	Hyodeoxycholic acid	Ginsenoside Rb2	0.0023791	0.5085757	2.799	0.00043235
191	Ginsenoside Rg3	Ginsenoside Rc	0.0071331	0.1689280	2.836	0.00042483
144	Ginsenoside Re	Cholic acid	0.0028026	0.3885680	2.605	0.00041806
150	Ginsenoside Re	Gamabufotalin	0.0025004	0.4493520	2.757	0.00040753
159	Ginsenoside Rg1	Bufalin	0.0026671	0.4468832	2.946	0.00040462
186	Ginsenoside Rg3	11-hydroxyprogesterone	0.0035909	0.3639052	3.288	0.00039744
143	Ginsenoside Re	Bufalin	0.0024996	0.4315939	2.746	0.00039280
175	Ginsenoside Rg1	Muscione	0.0027327	0.4093855	2.909	0.00038456
31	Borneol	Ginsenosde Rd	0.0050234	0.2286708	2.992	0.00038392
145	Ginsenoside Re	Chenodeoxycholic acid	0.0026935	0.3722801	2.647	0.00037876
228	Muscione	Ginsenoside Re	0.0025652	0.3803266	2.692	0.00036245
32	Borneol	Ginsenoside Re	0.0024919	0.4029976	2.781	0.00036113
194	Ginsenoside Rg3	Ginsenoside Rg1	0.0045883	0.2263224	2.880	0.00036056
158	Ginsenoside Rg1	Borneol	0.0026594	0.4035820	2.978	0.00036047
227	Muscione	Ginsenosde Rd	0.0050967	0.2056309	2.912	0.00035986
195	Ginsenoside Rg3	Hyodeoxycholic acid	0.0036520	0.2861029	2.922	0.00035763
148	Ginsenoside Re	Deoxycholic acid	0.0026924	0.3535186	2.676	0.00035564
161	Ginsenoside Rg1	Chenodeoxycholic acid	0.0028610	0.3504024	2.832	0.00035393
185	Ginsenoside Rg3	Deoxycholic acid	0.0036570	0.2722805	2.850	0.00034935
160	Ginsenoside Rg1	Cholic acid	0.0029701	0.3236896	2.766	0.00034758
165	Ginsenoside Rg1	Deoxycholic acid	0.0028598	0.3433353	2.827	0.00034727
187	Ginsenoside Rg3	Gamabufotalin	0.0034650	0.2985135	2.983	0.00034671
176	Ginsenoside Rg1	17-hydroxyprogesterone	0.0027929	0.3509641	2.828	0.00034664
147	Ginsenoside Re	Cinobufagin	0.0025995	0.3731483	2.799	0.00034655
174	Ginsenoside Rg1	Hyodeoxycholic acid	0.0028549	0.3525642	2.916	0.00034520
211	Hyodeoxycholic acid	Ginsenoside Re	0.0026874	0.3518738	2.744	0.00034461
101	Ginsenoside Rb2	Cinobufagin	0.0022913	0.4275393	2.879	0.00034021
193	Ginsenoside Rg3	Ginsenoside Re	0.0044208	0.2069379	2.692	0.00033979
181	Ginsenoside Rg3	Chenodeoxycholic acid	0.0036581	0.2612411	2.842	0.00033624
224	Muscione	Ginsenoside Rb2	0.0022570	0.4059781	2.734	0.00033510
156	Ginsenoside Re	17-hydroxyprogesterone	0.0026254	0.3388738	2.659	0.00033464
188	Ginsenoside Rg3	Ginsenoside Rb1	0.0074046	0.1253121	2.836	0.00032714
64	Cinnamaldehyde	Ginsenoside Re	0.0022887	0.3990170	2.796	0.00032660
179	Ginsenoside Rg3	Bufalin	0.0034642	0.2787548	2.966	0.00032553
63	Cinnamaldehyde	Ginsenosde Rd	0.0048201	0.2032004	3.015	0.00032488
162	Ginsenoside Rg1	Cinnamic acid	0.0023719	0.4170929	3.051	0.00032421
164	Ginsenoside Rg1	Cinobufagin	0.0027670	0.3495501	2.986	0.00032386
184	Ginsenoside Rg3	Cinobufagin	0.0035641	0.2727763	3.011	0.00032291

146	Ginsenoside Re	Cinnamic acid	0.0022044	0.4110862	2.831	0.00032012
163	Ginsenoside Rg1	Cinnamaldehyde	0.0024562	0.3813532	3.018	0.00031039
157	Ginsenoside Rg1	Benzyl benzoate	0.0025188	0.3799268	3.115	0.00030717
14	Benzyl benzoate	Ginsenoside Re	0.0023513	0.3734563	2.881	0.00030475
192	Ginsenoside Rg3	Ginsenoside Rd	0.0069522	0.1231611	2.874	0.00029794
10	Benzyl benzoate	Ginsenoside Rb2	0.0020430	0.4130322	2.850	0.00029609
166	Ginsenoside Rg1	11-hydroxyprogesterone	0.0027938	0.3452908	3.261	0.00029579
133	Ginsenoside Rd	Cinnamic acid	0.0047358	0.1872005	3.050	0.00029071
28	Borneol	Ginsenoside Rb2	0.0021837	0.3700235	2.805	0.00028809
149	Ginsenoside Re	11-hydroxyprogesterone	0.0026263	0.3263788	3.107	0.00027588
196	Ginsenoside Rg3	Muscione	0.0035298	0.2237633	2.902	0.00027215
13	Benzyl benzoate	Ginsenoside Rd	0.0048827	0.1716580	3.094	0.00027092
215	Muscione	Cholic acid	0.0019116	0.3942771	2.801	0.00026913
60	Cinnamaldehyde	Ginsenoside Rb2	0.0019805	0.3815768	2.831	0.00026697
97	Ginsenoside Rb2	Bufalin	0.0021914	0.3402093	2.819	0.00026450
104	Ginsenoside Rb2	Gamabufotalin	0.0021921	0.3271152	2.840	0.00025247
183	Ginsenoside Rg3	Cinnamaldehyde	0.0032533	0.2287672	2.994	0.00024855
178	Ginsenoside Rg3	Borneol	0.0034566	0.2123170	2.975	0.00024670
182	Ginsenoside Rg3	Cinnamic acid	0.0031690	0.2091304	3.034	0.00021844
2	Benzyl benzoate	Cholic acid	0.0016977	0.3797622	2.997	0.00021515
100	Ginsenoside Rb2	Cinnamic acid	0.0018962	0.3240599	2.860	0.00021482
177	Ginsenoside Rg3	Benzyl benzoate	0.0033159	0.1969754	3.069	0.00021280
81	Gamabufotalin	Cholic acid	0.0018468	0.3129027	2.828	0.00020431
41	Cholic acid	Cinnamic acid	0.0015508	0.3732538	2.933	0.00019734
220	Muscione	Deoxycholic acid	0.0018014	0.3025681	2.834	0.00019232
230	Muscione	17-hydroxyprogesterone	0.0017345	0.2999613	2.835	0.00018350
52	Cinnamaldehyde	Cholic acid	0.0016351	0.3230877	2.898	0.00018226
126	Ginsenoside Rc	Ginsenoside Rb1	0.0091523	0.0538366	2.710	0.00018184
39	Cholic acid	Bufalin	0.0018460	0.2688748	2.811	0.00017654
221	Muscione	11-hydroxyprogesterone	0.0017353	0.3254281	3.213	0.00017578
16	Benzyl benzoate	17-hydroxyprogesterone	0.0015205	0.3453838	3.036	0.00017301
216	Muscione	Chenodeoxycholic acid	0.0018025	0.2634202	2.810	0.00016898
229	Muscione	Hyodeoxycholic acid	0.0017964	0.2696328	2.874	0.00016856
19	Borneol	Cholic acid	0.0018384	0.2555726	2.834	0.00016579
7	Benzyl benzoate	11-hydroxyprogesterone	0.0015214	0.3449843	3.424	0.00015330
86	Gamabufotalin	11-hydroxyprogesterone	0.0016705	0.2856633	3.240	0.00014726
6	Benzyl benzoate	Deoxycholic acid	0.0015875	0.2811236	3.045	0.00014658
25	Borneol	11-hydroxyprogesterone	0.0016621	0.2826002	3.230	0.00014542
48	Cinnamic acid	11-hydroxyprogesterone	0.0013745	0.3524914	3.350	0.00014462
49	Cinnamic acid	17-hydroxyprogesterone	0.0013737	0.3122880	2.969	0.00014448
3	Benzyl benzoate	Chenodeoxycholic acid	0.0015886	0.2662406	3.023	0.00013989
68	Cinobufagin	Cholic acid	0.0019459	0.2060480	2.867	0.00013984
37	Bufalin	11-hydroxyprogesterone	0.0016697	0.2652003	3.230	0.00013710
85	Gamabufotalin	Deoxycholic acid	0.0017366	0.2227425	2.876	0.00013448
66	Cinnamaldehyde	17-hydroxyprogesterone	0.0014580	0.2701721	2.938	0.00013409
57	Cinnamaldehyde	11-hydroxyprogesterone	0.0014588	0.3016150	3.324	0.00013238
205	Hyodeoxycholic acid	Gamabufotalin	0.0017316	0.2155613	2.910	0.00012828
35	Borneol	17-hydroxyprogesterone	0.0016612	0.2209447	2.865	0.00012811

76	Deoxycholic acid	Cinnamic acid	0.0014406	0.2603647	2.974	0.00012613
82	Gamabufotalin	Chenodeoxycholic acid	0.0017377	0.2048424	2.856	0.00012462
15	Benzyl benzoate	Hyodeoxycholic acid	0.0015825	0.2425826	3.086	0.00012438
87	Gamabufotalin	17-hydroxyprogesterone	0.0016696	0.2087172	2.850	0.00012226
56	Cinnamaldehyde	Deoxycholic acid	0.0015249	0.2314236	2.940	0.00012001
38	Bufalin	17-hydroxyprogesterone	0.0016689	0.1953745	2.846	0.00011458
45	Chenodeoxycholic acid	Cinnamic acid	0.0014417	0.2343363	2.952	0.00011443
214	Muscione	Bufalin	0.0016086	0.1982825	2.845	0.00011212
73	Deoxycholic acid	Bufalin	0.0017358	0.1824320	2.864	0.00011058
222	Muscione	Gamabufotalin	0.0016094	0.1949796	2.851	0.00011008
201	Hyodeoxycholic acid	Cinnamic acid	0.0014356	0.2294984	3.015	0.00010927
53	Cinnamaldehyde	Chenodeoxycholic acid	0.0015260	0.2082235	2.917	0.00010894
219	Muscione	Cinobufagin	0.0017085	0.1852218	2.917	0.00010848
198	Hyodeoxycholic acid	Bufalin	0.0017308	0.1763738	2.901	0.00010524
24	Borneol	Deoxycholic acid	0.0017281	0.1737789	2.866	0.00010479
44	Chenodeoxycholic acid	Bufalin	0.0017369	0.1706277	2.835	0.00010453
65	Cinnamaldehyde	Hyodeoxycholic acid	0.0015199	0.1970686	2.982	0.00010043
8	Benzyl benzoate	Gamabufotalin	0.0013955	0.2207313	3.080	0.00010002
5	Benzyl benzoate	Cinobufagin	0.0014946	0.2093215	3.139	0.00009968
69	Cinobufagin	Chenodeoxycholic acid	0.0018368	0.1535312	2.902	0.00009719
46	Chenodeoxycholic acid	11-hydroxyprogesterone	0.0018636	0.1605815	3.162	0.00009464
33	Borneol	Hyodeoxycholic acid	0.0017231	0.1587758	2.911	0.00009400
1	Benzyl benzoate	Bufalin	0.0013947	0.2052891	3.075	0.00009311
77	Deoxycholic acid	Cinobufagin	0.0018357	0.1411650	2.919	0.00008878
71	Cinobufagin	11-hydroxyprogesterone	0.0017696	0.1627538	3.282	0.00008776
20	Borneol	Chenodeoxycholic acid	0.0017293	0.1432727	2.850	0.00008694
23	Borneol	Cinobufagin	0.0016353	0.1572838	2.960	0.00008690
202	Hyodeoxycholic acid	Cinobufagin	0.0018307	0.1390198	2.958	0.00008604
55	Cinnamaldehyde	Cinobufagin	0.0014320	0.1816085	3.035	0.00008569
72	Cinobufagin	17-hydroxyprogesterone	0.0017688	0.1364611	2.883	0.00008371
70	Cinobufagin	Cinnamic acid	0.0013477	0.1875296	3.070	0.00008233
47	Chenodeoxycholic acid	17-hydroxyprogesterone	0.0018628	0.1163772	2.764	0.00007842
212	Hyodeoxycholic acid	17-hydroxyprogesterone	0.0018566	0.1182700	2.830	0.00007760
34	Borneol	Muscione	0.0016010	0.1341892	2.856	0.00007522
204	Hyodeoxycholic acid	11-hydroxyprogesterone	0.0018575	0.1251134	3.229	0.00007197
43	Cholic acid	17-hydroxyprogesterone	0.0019719	0.0968127	2.709	0.00007047
42	Cholic acid	11-hydroxyprogesterone	0.0019727	0.1109740	3.117	0.00007023
79	Deoxycholic acid	17-hydroxyprogesterone	0.0018616	0.1010832	2.768	0.00006798
84	Gamabufotalin	Cinobufagin	0.0016437	0.1142128	2.939	0.00006387
199	Hyodeoxycholic acid	Cholic acid	0.0020338	0.0869326	2.796	0.00006324
78	Deoxycholic acid	11-hydroxyprogesterone	0.0018625	0.1039276	3.182	0.00006082
17	Borneol	Benzyl benzoate	0.0013870	0.1329471	3.087	0.00005974
58	Cinnamaldehyde	Gamabufotalin	0.0013329	0.1227309	2.978	0.00005492
67	Cinobufagin	Bufalin	0.0016429	0.0914737	2.930	0.00005129
83	Gamabufotalin	Cinnamic acid	0.0012486	0.1140554	3.010	0.00004731
51	Cinnamaldehyde	Bufalin	0.0013321	0.1032419	2.968	0.00004633
74	Deoxycholic acid	Cholic acid	0.0020388	0.0615986	2.724	0.00004611
40	Cholic acid	Chenodeoxycholic acid	0.0020399	0.0601707	2.721	0.00004511

218	Muscone	Cinnamaldehyde	0.0013977	0.0919643	2.918	0.00004405
36	Bufalin	Cinnamic acid	0.0012478	0.1024529	3.001	0.00004260
26	Borneol	Gamabufotalin	0.0015361	0.0794001	2.889	0.00004221
213	Muscone	Benzyl benzoate	0.0014603	0.0832178	3.034	0.00004005
217	Muscone	Cinnamic acid	0.0013134	0.0842688	2.956	0.00003744
18	Borneol	Bufalin	0.0015354	0.0682438	2.879	0.00003639
75	Deoxycholic acid	Chenodeoxycholic acid	0.0019297	0.0510757	2.783	0.00003541
231	17-hydroxyprogesterone	11-hydroxyprogesterone	0.0017956	0.0550709	3.191	0.00003099
200	Hyodeoxycholic acid	Chenodeoxycholic acid	0.0019247	0.0372885	2.849	0.00002519
50	Cinnamaldehyde	Benzyl benzoate	0.0011838	0.0627316	3.132	0.00002371
22	Borneol	Cinnamaldehyde	0.0013245	0.0479287	2.974	0.00002135
4	Benzyl benzoate	Cinnamic acid	0.0010995	0.0606778	3.167	0.00002107
21	Borneol	Cinnamic acid	0.0012402	0.0487210	3.004	0.00002011
203	Hyodeoxycholic acid	Deoxycholic acid	0.0019236	0.0277914	2.853	0.00001874
80	Gamabufotalin	Bufalin	0.0015438	0.0178793	2.873	0.00000961
54	Cinnamaldehyde	Cinnamic acid	0.0010369	0.0143999	3.065	0.00000487

Figure A4.1: The dose response matrices of five combinations from the 20 top-ranked combinations.

